UNIVERSITAT POLITECNICA DE CATALUNYA

MASTER'S THESIS

Clinical Assessment of Shock through Machine Learning Techniques

Author:
Alex AUSHEV

Supervisors:
Dr. Alfredo VELLIDO
Dr. Vicent RIBAS

A thesis submitted in fulfillment of the requirements for the degree of Master in Artificial Intelligence

in the

Facultat d'Informatica
de Barcelona

April, 2018

UNIVERSITAT POLITECNICA DE CATALUNYA

Abstract

Facultat d'Informatica
de Barcelona

Master in Artificial Intelligence

Clinical Assessment of Shock through Machine Learning Techniques

by Alex AUSHEV

Circulatory shock is a life-threatening disease that accounts for around one-third of all admissions to intensive care units (ICU). It requires immediate treatment and can result in irreversible damage to organs or even death. That is why it is important to provide fast and interpretable tools that can detect a clear therapeutic target for dealing with shock. This work uses machine learning techniques combined with clinical data to identify such targets.

This Master's thesis attempts to identify biomarkers that can be used as indicators for mortality prediction in patients with shock. This work uses the ShockOmics project dataset (75 observations and 333 features) to extract biomarkes that show good performance as features in the outcome prediction task. Multiple imputation techniques and machine learning models were used in the preliminary experiments in order to analyze the data set. Then, several feature selection techniques were applied to different subsets of the complete dataset. However, only the univariate feature selection based on ANOVA F-values and the Random Forest-based selection were able to achieve six promising feature sets. They were later used to build causal Bayesian Network structures that revealed some of the causal relationships between individual features and the outcome.

The main result of this work is a list of candidate features for mortality prediction. Some of the indicators in the list (e.g. *SOFA*, *APACHE II*) are well-known, and some of them (e.g. *Inferior vena cava distensibility index*, *X Norepinephrine*) require more thorough consideration. Moreover, the findings show that biomarkers have the most predictive value at certain time-steps. This means that, as shock progresses, different biomarkes should be prioritized. The main limitation of this study is the low amount of observations available for analysis. In future work, the findings of this thesis should be tested on a larger dataset.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Shock (or circulatory shock) is a life-threatening medical condition that requires immediate treatment. It occurs when the organs and tissues of the body are not receiving enough blood. As a result, the cells do not receive oxygen and nutrients from the blood, and the organs become damaged. Low blood pressure, fast pulse rate, rapid breathing and fever are among the common symptoms [1]. The outcome of an individual case depends on the stage of shock, the underlying condition and the general medical state of the patient [1]. It is crucial to detect this condition early since it can result in irreversible damage to organs or even death.

The mortality rate of shock remains very high and depends on its type. For example, it is around 30% for septic shock (according to the new definition of such condition [2]) and between 35.9% and 64.7% for in-hospital mortality in patients with cardiogenic shock (depending on the stage, [3]). Usually, shock is treated in the Intensive Care Unit (ICU), where it is carefully monitored. As a result, large amounts of data are produced for each patient. And hence, multiple omes (e.g. genome, proteome, transcriptome) sequencing can be used for identification of shock. Using such data is the essence of the multi-omics approach.

Machine learning (ML) techniques can help find the required biomarkers and predict the outcome of the patients with shock, using multi-omics data. Mult-omics approach requires a vast spectrum of features for each patient and, as a result, multi-omics data has very high dimensionality. Combining statistics and ML, one can retrieve new knowledge and causal relationships between different features from such high-dimensional data [4]–[6]. Having a model that may improve the general understanding of shock and help reduce the mortality rates can be a great asset in the ICU environment, where time is of the essence.

This work analyzes the dataset that was gathered within the ShocOmics research project. Usually, there are four types of shock that are commonly defined: hypovolaemic shock (e.g. hemorrhagic shock), cardiogenic shock, distributive shock (e.g. septic shock) and obstructive shock. However, in the collected dataset there are only two types of shock: septic and cardiogenic. Septic shock is the most common, and it is caused by infection. The second most frequent form of shock is the cardiogenic shock and it is a result of a heart failure. The multiscale nature of the ShockOmics dataset allows to search for biomarkers that are related to mortality induced by these two types of shock. This work attempts to analyze one part of the ShockOmics data, namely clinical data.

1.1 Motivation of the Thesis

When it comes to the detection of circulatory shock, physicians and therapists largely depend on a combination of clinical, hemodynamic and biochemical signs. For a treatment to be chosen, prompt identification of shock is necessary. Appropriate

treatment is based on a good understanding of the physiological mechanisms behind the condition. And shock is difficult to treat. Usually, there are other conditions that go along with shock (e.g. hypotension and hemodynamic instability, inflammation and multiple organ failure (MOF)). Since they have similar symptoms, it makes addressing causes of shock very difficult and hence, the therapist can not act properly at the beginning of the condition. For this reason, it is important to provide fast and interpretable tools that can detect a clear therapeutic target in order to prevent mortality or any irreversible consequences caused by shock.

Current research is focused on assessing the potential of different marker targets for resuscitation related to different stages of shock. Extensive description of the most common practices in shock diagnosis and treatment can be found in [7]. However, there are still many issues that can be addressed. For example, the potential of microcirculatory markers of tissue perfusion in microcirculation-oriented or microcirculation-guided management and/or therapy of early shock resuscitation [7] and the duration of hemodynamic effects of colloids and crystalloids after initial resuscitation [8]. Additionally, there are no simple and unambiguous clinical criteria or biological, imaging, or laboratory features that uniquely identify a septic patient [9], which makes the treatment of septic shock very difficult. Moreover, the research related to shock is compounded by several factors: the origin of shock, the degree of vasodilation, the supposed degree of capillary leak, and the shock severity. Current research will benefit from standardized clinical examination and the definition of appropriate statistical indexes. Clinical data combined with ML techniques may statistically estimate the therapeutic target for circulatory shock, providing the required analytical tools for studying and prognosis of shock. And can be later used as a part of the larger multi-omics study.

1.2 Thesis Objectives

The main objectives of this Master's thesis are set around identifying the biomarkers that can predict mortality of a patient with shock. This can help defining new targets for therapy in order to overcome the shortcomings of current therapies for circulatory shock. Along the way, several goals should be achieved, listed as follows:

1. *Studying evolution of shock in the context of these biomarkers.* It is important to analyze the biomarkers at different time steps. Because as shock progresses, it shows different symptoms, and these symptoms should be addressed with different priorities. The biomarkers that can predict mortality of a patient at different time steps can reveal the priorities for therapeutic targets and reduce any potential harm caused by shock.

2. *Identifying relationships between these biomarkers.* Addressing the importance of a single biomarker might not be enough. Depending on the whole set of biomarkers, individual biomarkers have different diagnostic values. For example, having all important aspects covered by a set of biomarkers is preferable than having a set of important biomarkers that covers only a few aspects. This is why it is important to study the influence of biomarkers in groups.

3. *Assessing effectiveness of already existing scoring systems for mortality prediction.* The current scoring systems (e.g. SOFA, APACHE II) are calculated at admission of a patient to the ICU. But there might be other variables that play an important role as indicators for mortality prognosis. Better understanding of

the mortality prediction value of existing scoring systems may help to identify what such systems are missing. This information can be later used for improving the mortality prediction models that are based on the extensive use of scoring systems.

4. *Building a model for ICU mortality predictions based on the ShockOmics dataset.* Finally, three previous points rely on an ML model that can predict mortality in the ShockOmics dataset. Trying out different sets of biomarkers (feature sets) and evaluating the model can provide us with an estimation of the importance of biomarkers. Naturally, we expect that the models that use important features should have a higher score. A model that can predict mortality is a valuable analytical tool for studying biomarkers and the evolution of shock in the ICU.

1.3 Considerations on the Analyzed Dataset

This Master's thesis analyses the dataset that was collected for the ShockOmics research project. According to [10]:

> The ShockOmics study is a multicenter prospective observational trial aimed at identifying new biomarkers of acute heart failure in circulatory shock, by means of a multiscale analysis of blood samples and hemodynamic data from subjects with circulatory shock.

The patients for the study were recruited in three ICUs: Hopital Erasme, Universite Libre de Bruxelles, Belgium; Hospital Universitari Mutua Terrassa, Spain; Hopitaux Universitaires de Geneve, Switzerland. This dataset allows to apply the multi-omics approach for analysis of clinical data. It has multiscale biomarkers from gene expression to protein synthesis, and from metabolite expression to organ specific injury. In this thesis only clinical data of the ShockOmics dataset is used.

The clinical data and blood samples were collected at three fixed time-steps: less than 16 hours, 48 hours and 7 days after the shock diagnosis. The dataset has 75 patients and more than 300 features that include different scores (more about scoring systems can be found in Chapter 2), analyses results and vital signs that are used for monitoring purposes. There are patients with septic shock, cardiogenic shock and a control group of patients that do not have shock. This dataset is described in detail in Chapter 3.

1.4 Expected Contributions

The main contribution of this Master's thesis can be viewed from two different viewpoints: clinical and ML-related. From a clinical point of view, it is expected to define the candidate biomarkers that can be successfully used as prognostic factors for the prediction of mortality due to shock. These biomarkers should be identified as novel targets for therapy to prevent shock progression. This can improve the overall accuracy of already existing prognostic tools and reduce any irreversible harmful impact on patients treated for circulatory shock.

From the ML point of view, it is expected to build a probabilistic graphical model that can predict mortality due to shock. The model should achieve high performance with the ShockOmics dataset and have a fixed feature set. This feature set needs to be a result of the thorough analysis of the initial dataset. For this reason, we expect

to perform multiple feature selection techniques and select the most promising features from that. The graphical model should incorporate the relations between the variables and predict the mortality, given full or partial values for those features. So analysis of the architecture of the this model is another contribution of the thesis.

1.5 Thesis Structure

The thesis is organized as follows:

- **Chapter 2** describes the condition of shock in more details. The chapter's main focus is to provide an overview of the pathology of shock: its symptoms, epidemiology and treatment. The emphasis is on septic and cardiogenic shock, since they are the most common types and since other types of shock are not present in the dataset. This chapter provides the medical background that is relevant to the problem of the thesis.

- **Chapter 3** focuses on the descriptions of the techniques that are currently used for mortality prediction and diagnosis of shock. This chapter introduces the multi-omics approach for analysis of diseases, scoring systems that are commonly used when shock is treated and the ML framework that is often used for the mortality prediction. The last section includes relevant work that shows different approaches to mortality prediction and analysis of biomarkers.

- **Chapter 4** describes preliminary experiments with the ShockOmics dataset. The experiments were done in the context of the analysis of the data. The chapter presents the description of the dataset along with the first results of mortality prediction. It describes multiple imputation techniques and ML models that were used to obtain the preliminary results. The best imputation technique and ML model are later used in the thesis.

- **Chapter 5** is dedicated to the feature selection procedure. It presents the results of the experiments with four subsets of the ShocOmics dataset and describes the feature selection techniques that were used. The purpose of these experiments is analyzing shock biomarkers. In the end of this chapter, several promising feature sets (biomarkers) are presented. They are later used in the remaining experiments.

- **Chapter 6** presents the set of the experiments, where different Bayesian Network structures are compared and analyzed. These structures provide the causal relationship information between biomarkers that allows identifying which features have the most influence on the outcome.

- **Chapter 7** presents the conclusions and suggestions for future work based on the results of the thesis.

Chapter 2

Medical Background: the Shock Pathology

As was mentioned in the previous chapter, circulatory shock is a life-threatening syndrome resulting in a multiorgan failure and a high mortality rate. It is best defined as a generalized form of acute circulatory failure associated with inadequate oxygen utilization by the cells. It is a state in which the circulation is unable to deliver sufficient oxygen to meet the demands of the tissues, resulting in cellular dysfunction [7]. In the 2014 *Consensus* on circulatory shock [7], the authors listed the quantitative and qualitative descriptions of indicators typically associated with shock. Patients with shock require a full clinical examination. It includes assessment of skin color and temperature, jugular venous distention, and peripheral edema. According to [1], the diagnosis can be refined with point-of-care echocardiographic evaluation, which includes assessment for pericardial effusion, measurement of left and right ventricular size and function, assessment for respiratory variations in vena cava dimensions, and calculation of the aortic velocity–time integral, a measure of stroke volume.

Circulatory shock can result from four mechanisms, described in [1]. The first of these mechanisms is a decrease in venous return due to a loss of circulating volume (i.e. due to internal or external loss of fluids). The second is a failure of the pump function of the heart that results from a loss of contractility (resulting from ischemia, infarction, myopathy, myocarditis), or a major arrhythmia (such as ventricular tachycardia, or a high degree A-V block). The third is an obstruction due to pulmonary embolism, tension pneumothorax or cardiac tamponade. The fourth is loss of vascular tone that results in maldistribution of blood flow (due to sepsis, anaphylaxis or spinal injury). Each of these mechanisms respectively correspond to four categories of shock: hypovolaemic, cardiogenic, obstructive and distributive. Some studies prove that a patient admitted with shock may later develop other types of shock [11], [12].

The rest of the chapter is built around the four categories of shock and their different aspects. The emphasis is made on septic and cardiogenic shock. They are particularly important for this thesis, since these are types of shock that are present in the ShockOmics dataset.

2.1 Diagnosis and Causes of Shock

A diagnosis of shock is based on clinical, hemodynamic, and biochemical signs. Circulatory shock is usually diagnosed by three main indicators that are typically present in the patients with shock [1]. The first one is systemic arterial hypotension. Usually in such patients, the systolic arterial pressure is less than 90 mm Hg or the

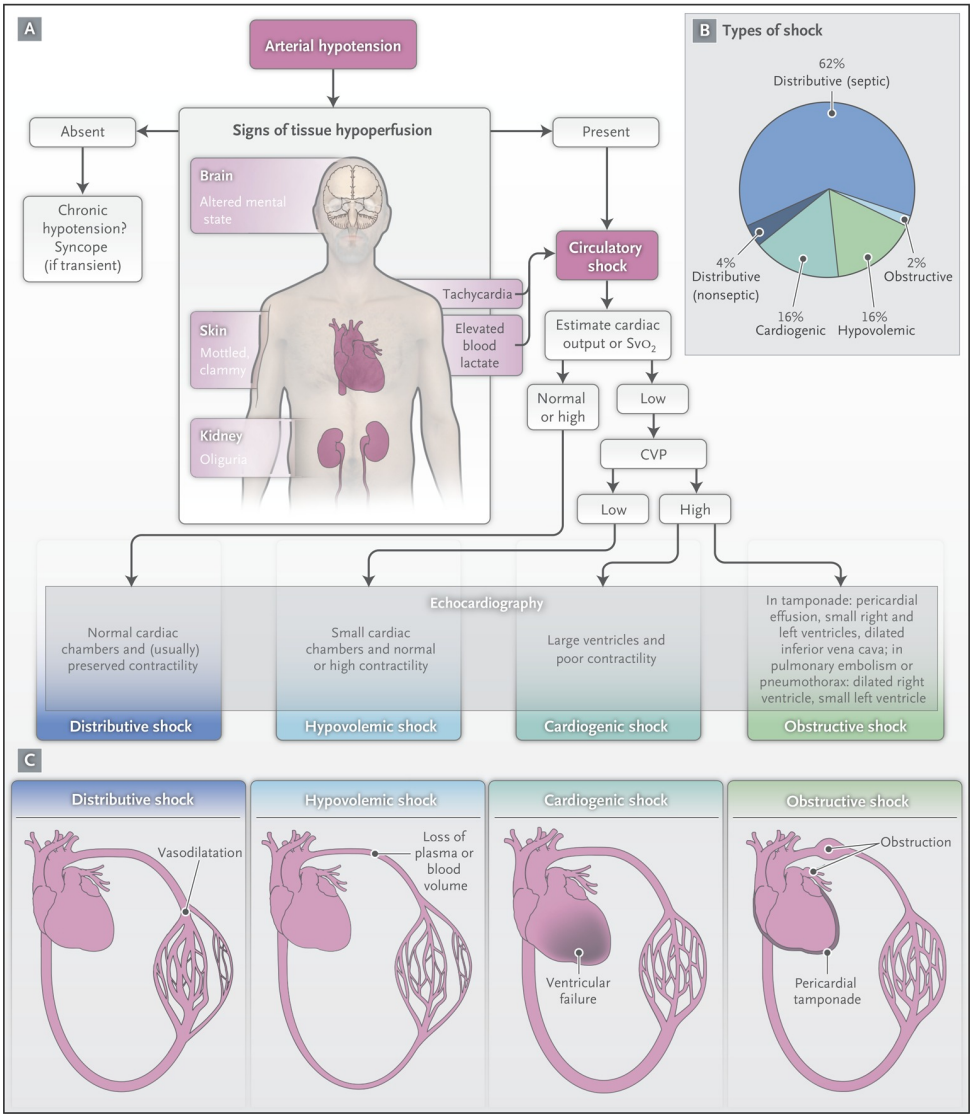


FIGURE 2.1: Initial assessment of shock states (source: [1]).

mean arterial pressure is less than 70 mm Hg. The second symptom is tissue hypoperfusion. It can be detected by cold and clammy skin, small urinal output and altered mental state. Lastly, hyperlactatemia is typically present in patients with circulatory shock, indicating abnormal cellular oxygen metabolism. These indicators are often revisited and they are a big part of the active research and discussion in the scientific community.

Having these symptoms present means that a patient has circulatory shock. However, this generic diagnosis is not enough and the specific type of shock should be identified as well. Figure 2.1 summarizes the pipeline for shock diagnosis that is presented in the rest of this section. It shows the initial assessment of a patient with shock (Panel A), relative frequencies of the main types of shock (Panel B), and schematic representations of the four main types of shock (Panel C) [1].

2.1.1 Distributive (Septic) Shock

The first step in the diagnosis pipeline is to eliminate the possibility of Distributive shock. For this reason, it is crucial to monitor the estimate cardiac output and SvO_2

(venous oxygen saturation). If one of these indicators is normal or high, it usually means that cardiac chambers are normal and heart contractility is preserved. And hence, distributive shock is diagnosed. Distributive shock occurs as a result of poor distribution of blood to the tissues, leading to inadequate tissue perfusion. Multiple types of shock fall under this category such as spinal, septic, and anaphylactic shock. The distributive shock is also known as relative hypovolaemia. It happens when an allergic reaction, or damage to the nervous system cause blood vessels to vasodilate or become leaky.

There are three subtypes of distributive shock: anaphylactic, neurogenic and septic shock. Anaphylactic shock refers to an allergic reaction that causes low blood pressure. Neurogenic shock is caused by a damaged central nervous system and also leads to low blood pressure. Septic shock is the most common type of distributive shock.

In the context of distributive shock, septic shock is of particular interest for this thesis. According to [13] and based on a consensus process using results from a systematic review, surveys, and cohort studies, septic shock is defined as a subset of the more general sepsis pathology in which underlying circulatory, cellular, and metabolic abnormalities are associated with a greater risk of mortality than sepsis alone. Adult patients with septic shock can be identified using the clinical criteria of hypotension requiring vasopressor therapy to maintain mean blood pressure of 65 *mm Hg* or greater and having a serum lactate level greater than 2 *mmol/L* after adequate fluid resuscitation.

2.1.2 Hypovolaemic (or Hemorrhagic) Shock

If the estimate cardiac output or SvO_2 is low, then the next step in the diagnostic pipeline is to check whether the shock is hypovolaemic. For this purpose, it is important to check the central venous pressure (CVP). If it is low the shock is hypovolaemic. This means that the contractility of a heart is high and that the cardiac chambers are small. This is a result from blood, plasma loss or excessive fluid loss (e.g., major burns). Usually, it requires more than 20 percent (one-fifth) of the body's blood or fluid supply to be lost. Hypovolaemic shock is very common with young children and older adults. It often results from a bleeding due to trauma (e.g. serious cuts, blunt injuries, internal bleeding, etc.). Among other causes of this type of shock, we find severe burns, excessive diarrhea, vomiting and sweating. The extensive loss fluid related to hypovolaemic shock causes a wide range of symptoms: from mild symptoms like headaches and nausea, to severe symptoms like pale skin, rapid breathing, rapid heart rate, lightheadedness and weak pulse. However, these symptoms are very common among other types of shock, so it is important to check for any types of wounds, internal bleeding and other indications of blood loss to detect hypovolaemic shock. Hypovolaemic shock occurs when dehydration or hemorrhage reduces blood volume.

2.1.3 Cardiogenic Shock

Once the hypovolaemic and cardiogenic shock diagnoses are discarded, there are still two candidate types of shock: cardiogenic or obstructive. Cardiogenic shock occurs when the heart is unable to circulate enough blood volume to maintain adequate tissue perfusion. It results in large ventricles and poor contractility due to a damage to the heart. Cardiogenic shock is a result of an acute episode of heart failure or a heart attack. In [14], the authors pointed out that cardiogenic shock has

clinical signs of hypoperfusion and/or serum lactate > 2 *mmol/L* despite fluid resuscitation (n = 219, mean age 67, 74% men). Usually, risks for cardiogenic shock include previous history of myocardial infarction (heart attack), plaque buildup in the coronary arteries (arteries supplying blood to the heart) and long-term valvular disease (disease affecting the valves of the heart).

Since most symptoms of cardiogenic shock are very similar to those of other types of shock, it is important to look for potential causes of a heart failure. They include inflammation of the heart muscle (myocarditis), infection of the heart valves (endocarditis), inability of the heart muscle to work properly (e.g. sudden blockage of an artery in the lung, damage to the valves allowing the backflow of blood, fluid buildup around the heart reducing its filling capacity, etc.), drug overdoses or poisoning with substances that can affect heart's pumping ability, etc. All these factors may damage the main pumping chamber of the heart and cause cardiogenic shock. Cardiogenic shock happens when injury prevents heart from pumping efficiently.

2.1.4 Obstructive Shock

If the estimate cardiac output or SvO_2 is low and CVP is high, then there is a possibility of obstructive shock. It can occur due to some kind of obstruction in the cardiovascular system like pulmonary embolism (blood clot in the lung), cardiac tamponade (compression of the heart due to fluid build up), tension pneumothorax (collapsed lung), heart lesions (obstruct the flow of blood from the heart), or vena cava syndrome (a major vein in the body becomes blocked and cannot carry blood from the body to the heart) [1]. This means that despite normal intravascular volume and myocardial function (heart pumping well), there are physical changes that lead to obstruction during diastolic filling of the ventricles, in contrast to biological or chemical changes. Hence, the flow of the blood to the heart is decreased due to obstruction. Signs, symptoms, management and treatment of obstructive shock are dependent upon the cause.

Signs and symptoms of obstructive shock are very similar to those of cardiogenic shock. The main difference is that obstructive shock is caused by an obstruction of blood flow outside of the heart. While in cardiogenic shock the cause of the shock is in the heart itself. Obstructive shock occurs due to a reduction in venous return, but may also be caused by blockage of the aorta.

2.2 Epidemiology of Shock

According to [15], up to one-third of patients admitted to the ICU are in circulatory shock. In the 1,679 ICU patients in the European Sepsis Occurrence in Acutely Ill Patients II trial, septic shock was the most frequent cause of shock, accounting for 62% of cases, followed by cardiogenic shock (16%), hypovolemic shock (in 16%), other types of distributive shock (in 4%), and obstructive shock (in 2%) [16]. Figure 2.1(B) shows relative frequencies of the main types of shock. Since the septic shock and cardiogenic shock are the most frequent and of particular interest for this thesis, they require more attention.

Septic shock is the most severe manifestation of sepsis. Based on the 1,656 patient study[16], the reported case-fatality rates due to septic shock are in the range of 40–50%, reaching as high as 80%. More recent studies report that the septic shock–associated crude mortality is 46.5% [13] and 58.8% [17] (see Figure 2.2). This systematic review identified 44 studies reporting septic shock outcomes (total of 166

	2007	2008	2009	2010	2011	2012	2013
Septic shock (R57.2)							
Cases	–	–	–	22 326	27 151	30 688	33 815
Deaths	–	–	–	13 616	16 143	18 024	19 891
Adjusted rate per 100 000 persons	–	–	–	27	33	37	40
In-hospital mortality (%)	–	–	–	61.0	59.5	58.7	58.8

FIGURE 2.2: Hospital incidence and mortality rates for septic shock in Germany, 2007-2013 (source: [17]).

479 patients) from a total of 92 sepsis epidemiology studies. The epidemiology of septic shock in low-income countries is very limited, although some literature suggest that its incidence is increasing [7]. Among the patients admitted to IC,U the reported incidence of septic shock is between 6.3% and 14.7% [18], [19]. The epidemiology of shock is complicated by differences in shock identification. Multiple studies report different cutoffs and combinations for blood pressure (BP), fluid resuscitation, vasopressors, serum lactate level, and base deficit to identify septic shock [7]. Septic shock is the most common form of shock and it remains the leading cause of mortality in the ICU.

Cardiogenic shock is less frequent compared to septic shock. It remains the second most frequent form of shock, though. It has most commonly been studied in the setting of acute myocardial infarction. Among these patients, 6-9% develop cardiogenic shock [12]. In a multinational observational study of 65,119 patients hospitalized for an acute coronary syndrome between 1999 and 2007, 4.6% developed cardiogenic shock, and the in-hospital case-fatality rate was 59.4% [7]. Cardiogenic shock is the leading cause of mortality in patients hospitalized with acute coronary syndrome (ACS). In a study conducted between 2001 and 2014 [20], among the 28,217 patients with ACS, 1,209 (4.3%) had cardiogenic shock: 526 (44%) at the time of admission and 683 (56%) later on during hospitalization. Among all age groups, a marked increase in the use of coronary angiography and revascularization, with a substantial reduction in the adjusted in-hospital mortality rate, was observed [20]. Overall in-hospital cardiogenic shock mortality decreased from 60.3% in 1995 to 47.9% in 2004 (P<.001) [12].

2.3 Treatment of Shock

Treatment of shock depends on a cause. It requires a good understanding of the underlying pathophysiological mechanisms and it is critical to identify the type of shock correctly. Even though the treatment varies depending on the shock type, there are general guidelines that are universal for all types. Treatment should be based on the correction of the cause of shock and hemodynamic stabilization through fluid infusion and administration of vasoactive agents. Blood lactate measurements and clinical evaluation are used for the patient's response monitoring [1].

According to [1], there are four phases in the treatment of shock (see Figure 2.3). During the first (salvage) phase, the goal is to stabilize the blood pressure and cardiac output that are required for survival. This phase requires performing lifesaving procedures to treat the underlying cause of shock. In the second (optimization) phase, the treatment is focused on increasing cellular oxygen availability and targeting hemodynamic status [21]. During this phase, cardiac output, mixed venous oxygen saturation (SvO_2), and lactate levels are carefully monitored. When the hemodynamic stability is achieved, it is time for the third (stabilization) phase,

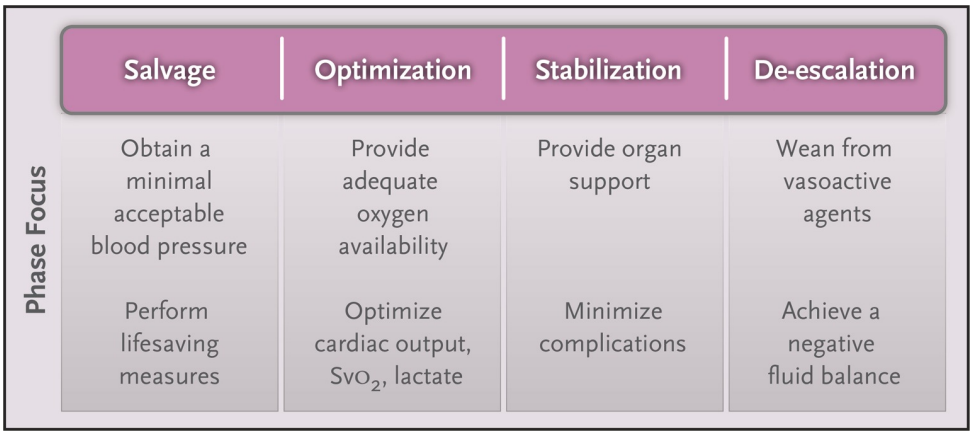


FIGURE 2.3: Four Phases in the Treatment of Shock (source: [1]).

where the goal is to prevent organ dysfunction, even after hemodynamic stability has been achieved. Finally, the fourth (de-escalation) phase attempts to wean the patient from vasoactive agents and promote spontaneous polyuria or provoke fluid elimination through the use of diuretics or ultrafiltration to achieve a negative fluid balance [1].

Resuscitation of the patient should be started in parallel with the investigation of the cause. The important components of resuscitation are summarized in the VIP rule: *ventilate* (oxygen administration), *infuse* (fluid resuscitation), and *pump* (administration of vasoactive agents)[1]. Once identified, the cause must be corrected.

Hemodynamic support restores blood pressure, provides adequate cellular metabolism and decreases the blood lactate level. Early, adequate hemodynamic support of patients in shock is crucial to prevent worsening organ dysfunction and failure [1]. In patients with shock and a blood lactate level of more than 3 *mmol/L*, Jansen *et al.* [22] found that targeting a decrease of at least 20% in the blood lactate level over a 2-hour period seemed to be associated with reduced in-hospital mortality. Hemodynamic support is crucial when it comes to prevention of worsening organ dysfunction and failure.

The four phases described in this section are essential for the circulatory shock treatment. The treatment should be immediate, while the cause of shock is being identified. The most important steps of the treatment are the stabilization of blood pressure so that heart and brain are perfused with blood, treatment of the cause of shock, increasing heart contractility, and providing oxygen and airway protection. Once the treatment has gone through all four phases, the patient's life is not in danger and the possible damage caused by shock is reduced.

Chapter 3

State of the Art: Mortality Prediction in the ICU

This chapter describes important aspects of the work related to this thesis. Even though the thesis uses only clinical data, it is a part of the ShockOmics research project the manages multi-omics data. To our knowledge, there is no multi-omics study of mortality prediction in patients with circulatory shock. But since the mortality prediction model for one medical condition can be extended to other diseases, the state of the art techniques for mortality prediction not related to circulatory shock are relevant for this thesis. Additionally, the chapter includes the description of the multi-omics approach, the benefits and challenges of using it, and the description of some common scoring systems that are often used as features in the machine learning (ML) framework. This chapter is built around the ML framework, the multi-omics data it can use, and scoring systems that incorporate relevant information for mortality prediction knowledge, reducing the amount of features in the model.

3.1 The Multi-omics Approach for Disease Analysis

In recent years, advances in omics technologies have allowed to gather a big amount of diverse data at different levels, including genomics, proteomics, peptidomics, transcriptomics and metabolomics. This has created an opportunity to improve the understanding of multiple medical conditions and the biological processes underneath them. Encompassing multiple omics levels in the data and using these data for analysis is often called a *multi-omics approach*. The analysis of such data has been used to build comprehensive and dynamic models for such difficult conditions as cancer, where it had a higher performance in clinical outcome prediction and higher stability [23], and chronic kidney disease, where it allowed to track molecular changes for the purpose of drug discovery [24]. Multi-omics models have been shown to substantially increase the confidence in the identified biomarkers, therapeutic targets and outcome prediction in multiple recent studies [23]–[27].

A pilot study for mortality prediction in patients with severe septic shock by Ferrario et al. [28] attempted to verify a metabolomic approach to determine changes in circulating metabolites able to characterize the progress of septic shock condition and to reveal the involved pathways for the ShockOmics project. In this study, they concluded that the use of omics tools able of examining physiological responses at system level is a particularly promising approach for complex and heterogeneous conditions such as septic shock. They also noted that it should be proved in a larger cohort by including different phenotypes (not only the severe patients) and multi-markers information such as multi-omics data [28].

This thesis addresses some of the proposals for future work of the Ferrario *et al*. study [28] by using multi-level biomarkers.

3.2 Scoring Systems

Scoring systems are widely used for mortality prediction in the ICU. Various scores are designed to determine the current health status of the patient, to increase understanding of the effectiveness of treatment and to optimize the use of hospital resources. Different scores are developed for certain conditions such as, for instance, cancer, liver failure, or sepsis. Scores are calculated from the data collected on the first day of stay in the ICU, throughout ICU stay, or for the first several days in the ICU. These calculations are based on physiological variables and medical history information. Many studies have shown the effectiveness of scoring systems in predicting hospital mortality and most of the available scores are comparable in terms of outcome prediction [29].

There are two types of scores: subjective and objective. Subjective scores are established by a panel of experts who choose the variables and assign a weight to each variable based on their personal expert opinion. The more abnormal the result, the higher the weight that is assigned [30]. Objective score variables are collected using logistic regression modeling techniques and clinical judgment to determine ranges and to assign weights [31]. A scoring system usually comprises two parts – a score (a number assigned to disease severity) and a probability model (equation giving the probability of hospital death of the patients) [31]. A model refines the ability of scores or scales to be used in comparing various groups of patients for the purpose of treatment, triage or comparative analysis, and thus helps in decision making [32].

Scoring systems are also used for the mortality prediction for patients with shock. Some of the scores are developed specifically for certain types of shock. For example, the CardShock risk Score [14] is used for patients with cardiogenic shock; mortality prediction performance for this score is 0.85 (area under the receiver operating characteristic curve).

In this section, three scoring systems are considered in more detail, since they are relevant to the ShockOmics data and important for this thesis: the sequential organ failure assessment score (SOFA); the acute physiology and chronic health evaluation II (APACHE II); and the Glasgow Coma Scale.

3.2.1 SOFA

The SOFA score (or sepsis-related organ failure assessment) was designed to track a status of the patient during the stay in an ICU over time. It is a subjective score with the data collected throughout the ICU stay or for the first three days [31]. Its main goal is to determine the extent of patient's organ failure. It is based on the degree of organ dysfunction data on six organ failures and is scored on a scale of 0-4 [31]. The guideline for calculating the SOFA score is shown in Figure 3.1. Although the score's main purpose is to assess organ dysfunction, it is reported to be a good indicator of prognosis during the first few days of ICU admission (both the mean and highest SOFA scores are particularly useful) [35].

The performance of the models that use SOFA for outcome prediction is usually assessed based on area under the receiver operating characteristic curve (AUC). In research that studied critically ill patients [29], the SOFA score showed an AUC

Organ system	Score					
	Variable	0	1	2	3	4
Pulmonary	Lowest PaO_2 (Torr)/FiO_2 (%)	>400	≤400	≤300	≤200+respiratory support	≤100+respiratory support
Coagulation	Lowest platelet (10^3/mm^3)	>150	≤150	≤100	≤50	≤20
Hepatic	Highest bilirubin (μmol/L)	<20	20-32	33-101	102-204	>204
Circulatory	Blood pressure status	Mean arterial pressure (mmHg) >70	Mean arterial pressure (mmHg) <70	Dopamine* dose≤5 or dobutamine any dose	Dopamine dose>5 or epinephrine≤0.1 or norepinephrine≤0.1	Dopamine dose>15 or epinephrine>0.1 or norepinephrine>0.1
Neurologic	GCS	15	13–14	10-12	6–9	<6
Renal	Highest creatinine level (μmol/L)	<110	110-170	171-299	300-440	>440
	Total urine output (mL/24 h)				<500	<200
Score	0-6	7-9	10-12	13-14	15	15-24
Score %	<10	15-20	15-20	50-60	>80	>90

PaO_2: (Torr) arterial oxygen tension; FiO_2: Fractional concentration of inspired oxygen; GCS: Glasgow coma score

FIGURE 3.1: Sequential organ failure assessment score [33]–[35] (source: [31]).

of 0.776 for day-1 scores. A more recent study [36] involving 184,875 patients supports this performance showing an AUC of 0.753. Moreover, Rivera-Fernández *et al*. [37] showed that 28-day mortality was related to mean and maximum daily SOFA scores in a cohort of patients who were critically ill with an AUC of 0.95. The SOFA score shows a significant prognostic value for in-hospital mortality prediction, outperforming SIRS criteria and the qSOFA score in an ICU setting [36].

3.2.2 APACHE II

APACHE II is a subjective score that was designed to measure the severity of disease for patients admitted to ICU. The score is calculated based on the data collected on the first ICU day [31]. It uses such variables as age, preexisting diseases, and 12 acute physiological variables to calculate a score from 0 to 71 [38]. The guideline for calculating the APACHE II score is presented in Figure 3.2. The focus on the disease severity measurement makes the APACHE II score a good mortality prediction tool.

APACHE II is often used for outcome prediction. According to [39], APACHE II scores were able to predict the mortality and correlated positively with actual mortality ($r = 0.84$, $p < 0.01$). These findings were supported by the work of Agarwal *et al*. [40], where the mean APACHE II scores correlated well with the surgical outcome in patients of perforation peritonitis and with the hospital and ICU stay. The mean Apache II score in survivors was 4.05 and in non-survivors was 12.75 and p value < 0.0001 which is highly significant [40].

Additionally, APACHE II scores of discharge and admission were very useful for predicting post-ICU mortality and ICU readmission during the same hospitalisation in a surgical ICU [41]. Based on their results, the discharge APACHE II score is confirmed to be useful in predicting post-ICU mortality and is superior to the admission APACHE II score in predicting early ICU readmission in surgical ICU patients [41]. The discharge and admission scores in predicting in-hospital mortality was 0.631 and 0.669, respectively; and 0.606 and 0.574 for predicting all forms of readmission [41]. The APACHE II score showed very promising results for outcome prediction, and combining it with other scores (e.g. IMPACT [42]) seems to improve the predictive accuracy.

3.2.3 Glasgow Coma Scale

The Glasgow Coma Scale (GCS) is rarely used for mortality prediction but it is often a part of other scoring systems. It is a part of both SOFA and APACHE II scores, as can be seen in Figure 3.1 and 3.2. The reason behind this is that the GCS was

A: Acute physiological score (12 variables)

Physiologic variable	High abnormal range				Normal range	Low abnormal range			
	+4	+3	+2	+1	0	+1	+2	+3	+4
Temperature rectal (°C)	≥41	39-40.9	-	38.5-38.9	36-38.4	34-35.9	32-33.9	30-31.9	≤29.0
Mean arterial pressure (mm Hg)	≥160	130-159	110-129		70-109		50-69		≤49
Heart rate-ventricular response	≥180	140-179	110-139		70-109		55-69	40-54	≤39
Respiratory rate per minute-non-ventilated or ventilated	≥50	35-490		25-34	12-24	10-11	6-9		≤5
Oxygen: A-a DO_2 or PaO_2 (Torr)									
FiO_2 ≥0.5 record A-a DO_2	≥500	350-499	200-349		≤200	PO_2 61-70		PO_2 55-60	PO_2 <55
FiO_2 <0.5 record only PaO_2					PO_2 >70				
Arterial pH	≥7.7	7.6-7.69		7.5-7.59	7.33-7.49		7.25-7.32	7.15-7.24	<7.15
Serum HCO_3 (mmol/L)-only if no ABGs	≥52	41-51.9		32-40.9	23-31.9		18-21.9	15-17.9	<15
Serum sodium (mmoL/L)	≥180	160-179	155-159	150-154	130-149		120-129	111-119	≤110
Serum potassium (mmoL/L)	≥7	6-6.9		5.5-5.9	3.5-5.4	3-3.4	2.5-2.9		≤2.5
Serum creatinine (μmoL/L)	≥350	200-340	150-190		60-140		<60		
Hematocrit (%)	≥60		50-50.9	46-49.9	30-45.9		20-29.9		≤20
White blood cell count (×1,000/mm³)	≥40		20-39.9	15-19.9	3-14.9		1-2.9		<1
Glasgow coma score=15 minus actual GCS									

B: Age points		C: Chronic health points			Apache II score
Age (years)	Points	History	Points for elective surgery	Points for emergency surgery	Sum of A+B+C
≤44	0	Liver: Biopsy-proven cirrhosis and documented portal hypertension or prior episodes of hepatic failure	2	5	A: APS
45-54	2	Cardiovascular: NYHA Class IV	2	5	B: Age points
55-64	3	Respiratory: e.g., severe COPD, hypercapnia, home O_2, pulmonary hypertension	2	5	score
65-74	5	Immunocompromised	2	5	C: Chronic health
≥75	6	Renal: Chronic dialysis	2	5	point score
Total score					

APACHE: Acute physiology and chronic health evaluation; A-a DO_2: Alveolar-arterial oxygen tension difference; PaO_2 (Torr) arterial oxygen tension; FiO_2 (%): Fractional concentration of inspired oxygen; HCO_3: Bicarbonate; ABG: Arterial blood gas; NYHA: New York heart association; COPD: Chronic obstructive pulmonary disease. To compute predicted death rates for groups of acutely ill patients, the individual risk of hospital death is calculated with the following equation; the individual risks are then summed up and the value is divided by the total number of patients. $R/1\text{-}R$=−3.517+(APACHE II score×0.146)+(0.603, only if post-emergency surgery)+(diagnostic category weight as shown below), where R is the estimated risk of hospital death

FIGURE 3.2: Acute physiologic and chronic health evaluation (APACHE II) [38] (source: [31]).

designed for assessing the depth and duration of impaired consciousness and coma [43]. And as a result it provides the status of the central nervous system. Several studies have shown that there is a good correlation between GCS and neurological outcome [31], [44]. Currently, GCS is often used in studies related to traumatic brain injury mortality [45], [46]. Despite the focus of GCS on central nervous system and recommendations to apply it in the context of other physiologic information and the patient's specific diagnosis [47], it is still useful for mortality prediction. The guideline for calculating GCS is presented in Figure 3.3.

GCS was successfully used for mortality prediction in a big study involving 1,695 critically ill patients [48], it yielded an AUROC of 0.715. Higher results were shown in an earlier study [49], where the outcome was predicted among comatose patients. In this study, GCS yielded an AUROC of 0.87. This considerable difference in the performance between the two studies can be explained by the fact that GCS was primarily designed for patients in coma. Other studies also report the significance of GCS for mortality prediction (e.g. for organophosphate poisoning [50]), however, using GCS alone for this task is less common than using SOFA or APACHE II.

3.3 Machine Learning Framework

In the ICU, it is often difficult to accurately predict when patients develop shock or other critical conditions occur. As was mentioned, shock can be detected only afterwards, as most of the critical conditions that require immediate treatment. And at that time the damage (e.g organ injury) is already present. Multiple studies proved that critical physiological changes are seen in 51–86% of patients who suffered a subsequent cardiopulmonary arrest on general wards, often several hours before the

Score	Best eye response (E)
1	No eye opening
2	Eye opening to pain
3	Eye opening to verbal command
4	Eyes open spontaneously
Score	**Best verbal response (V)**
1	No verbal response
2	Incomprehensible sounds
3	Inappropriate words
4	Confused
5	Oriented
Score	**Best motor response (M)**
1	No motor response
2	Extension to pain
3	Flexion to pain
4	Withdrawal from pain
5	Localizing pain
6	Obeys commands

A coma score of 13 or higher correlates with a mild brain injury, 9-12 is a moderate injury and 8 or less a severe brain injury

The pediatric Glasgow coma scale[41,42]

Score	Eye opening	Verbal response	Motor response
6	-	-	Normal, spontaneous, obeys commands
5	-	Age-appropriate words, social smile, fixes and follows	Localizes pain
4	Spontaneous	Cries, but consolable	Withdraws from pain
3	To voice	Persistently irritable	Flexion posture to pain
2	To pain	Restless agitated	Extension posture to pain
1	None	None	None

FIGURE 3.3: Glasgow coma score [43] (source: [31]).

arrest event [51]. And if some critical conditions can be predicted by tracking physiologic changes, then there is a need to identify which variables should be monitored. This is not a trivial task.

Predicting the occurrence of some critical condition (e.g. circulatory shock) requires a minimal monitoring dataset to identify patients with a targeted condition across all possible processes, and then specifically monitor their responses to targeted therapies known to improve outcome [52]. For this problem, an ML framework can be used. Since prediction of outcome is a classification problem, then it requires multivariable datasets and data-driven classification techniques. Usually, this approach involves libraries of responses across large and comprehensive collections of records of diverse subjects whose diagnosis, therapies, and course is already known to predict not only disease severity, but also the subsequent behavior of the subject if left untreated or treated with one of the many therapeutic options [52]. This approach is beyond human intellectual scope and requires more advanced and data-driven techniques, such as ML methods.

Various ML algorithms can be used for pattern detection in multi-variable data for outcome prediction or qualitative analysis of medical conditions. Over the last decade, they have gained a lot of attention, since they have shown better performance than traditional statistical approaches, work well with large amounts of data and can be used in various tasks that occur in the ICU: from outcome prediction to therapeutic targets detection. And an improved detection of the physiological changes, together with the identification of new biomarkers (that a human observer can easily miss), is crucial in impacting the potential to rescue unstable patients.

3.3.1 Comparison with Statistical Approaches

In statistical approaches, logistic regression is commonly used for outcome prediction. However, even though this approach is very common, it is less accurate compared to ML techniques. There are multiple studies that support this statement and, as a result, more studies use an ML framework even in cases where single scores can be used.

ML techniques can outperform scoring systems and logistic regression for mortality prediction. In a recent study, Motwani *et al.* [53] compared LogitBoost, a boosted ensemble algorithm, with existing clinical metrics (FRS, SSS, SIS, DI) and logistic regression. They used coronary computed tomographic angiography (CCTA) data to evaluate prognostic risk in individual patients with suspected coronary artery disease. Their findings suggest that ML algorithms predicted 5-year ACM significantly better than existing clinical or CCTA metrics alone (AUC): 0.79 vs. FRS: 0.61, SSS: 0.64, SIS: 0.64, DI: 0.62; $P < 0.001$ [53]. Another study by Allyn *et al.* [54] provided similar results, comparing a greedy ensemble algorithm (a weighted sum of Gradient Boosting Machine, Random Forest, Support Vector Machine and Naive Bayes) with traditional methods for predicting mortality after elective cardiac surgery. AUROC for the ML model was 0.795 against 0.737 for the scoring system and 0.742 for logistic regression. Finally, the ML model (Random Forest) outperformed the Modified Early Warning score and logistic regression models in predicting clinical deterioration on the wards (AUC): 0.8 *vs.* 0.7 and 0.77 [55].

The ML framework effectively individualizes risk assessment and overcomes many of the limitations of a standard statistical approach. ML models are more accurate in predicting mortality and it justifies their usage in the field of medical prediction [55]–[57]. In the work of Taylor *et al.* [56], relevant for this thesis, they

compared ML and logistic regression models for mortality prediction of 4,676 patients with sepsis. Once again, based on AUC, logistic regression was outperformed (0.86 *vs.* 0.76) [56]. However, in predicting 30-day Heart Failure readmission, the ML algorithms (a tree-augmented naive Bayesian network, a random forest algorithm, and a gradient-boosted model) did not improve the results compared to traditional statistical methods using two independently derived logistic regression models and a least absolute shrinkage and selection operator method [58].

3.3.2 Multi-Omics Data Usage

The ML methodology scales up correlational analyses to potentially very highly multivariate, high-frequency, and perhaps multisource data that could help empirically discover leading indicators of instability [52]. Integrating multi-omics data into an ML framework has been a focus of much attention in the past years. Before that, little has been done to use multiple levels of omics biomarkers for improving the prediction accuracy. While prediction modeling using a single type of omics markers is a well-studied topic, it is not clear how different types of biomarkers should be handled simultaneously when deriving a prediction model [25]. But encompassing biomarkers on multiple levels might be a difficult task, since it requires consideration of hundreds or even thousands variables. And it is much more difficult to implement in clinical practice than a model including only a handful of variables [25].

Models including several hundreds/thousands of variables raise a problem of choosing an 'optimal' subset of variables for prediction, which is a critical task in personalized medicine. Using a large number of variables in ICU is not practical, since it requires a lot of monitoring efforts and cost. Unfortunately, although using omic markers for prediction has been a well-studied topic, it is not clear how the different modalities should be handled [59]. In general, there are two common approaches on how to handle different levels and modalities. The first one is to simply merge all the data and ignore the source of the variables. The second one is more sophisticated, and it requires analyzing each modality on its own and then merging the results [60]. Some authors even suggest that merging can be performed at different stages of the analysis [61]. Even though the literature is often vague on when to use different strategies [59], one should always consider using a small number of variables, as opposed to a large number of variables.

Numerous studies involve an ML framework for predicting outcome with multi-omics data. Some of them use traditional ML methods like Random Forests [62] to provide new insights, for example, into the multifactorial nature of Crohn's disease and help highlight cohort-specific to disease pathophysiology [26]. The model that Douglas *et al*. [26] presented in their study could be extended in the future to other diseases and to other data types such as transcriptomics and metabolomics to better understand the relative importance of each of these features.

Other studies tried to provide a novel systematic approaches to identify interaction models between pathways based on multi-omics data. For example, different modifications of the lasso method [63], such as priority-Lasso [25] and IPF-LASSO [59], were successfully applied recently. Different evolutionary algorithms are also used for mortality prediction with multi-omics data. The Grammatical Evolution Neural Networks method is proposed by Kim and colleagues [27] in their metadimensional knowledge-driven genomic interactions study. They claim that their novel framework identifies interaction models between pathways based on multi-omics data and incorporates biological knowledge into such models [27]. In a

different study by Kim *et al*. [64] an expression balance analysis algorithm was used to predict expression in a gene knockout condition, and the predictive performance of their model ranges from 0.54 to 0.87 for the various omics layers, which far exceeds various baselines [64]. They drew a conclusion that determining variability in molecular signatures based on these interactions between pathways may lead to better diagnostic/treatment strategies for better precision medicine [27].

3.3.3 Other Related Work

Various ML algorithms have successfully been used for outcome prediction. The alternating decision tree model showed an accurate continuous probabilistic score for the prediction of day 100 overall transplantation-related mortality [65]. Multiple methods like Neural Networks, Naive Bayes, Decision Trees, and logistic regression were compared in traumatic brain injury outcome prediction: Neural Networks was the best for 6-month functional outcome prediction (AUROC of 0.961), and Naive Bayes was the best predictive model for mortality (AUROC of 0.911) [66]. Probabilistic co-occurrence analysis and principal component analysis were applied in a leading indicators of adverse health events study, the AUROC of the proposed framework was 0.857 [67]. Probabilistic co-occurrence analysis was used for to segment each high-frequency measurement channel into sequences of non-overlapping time intervals, and to independently characterize waveform frequency spectrum in each segment, while principal component analysis applied to the spectra observed during a specific physiologic state (e.g., periods of stability) to envelope the range of variability of waveform patterns that can be expected in the particular state [67].

Some studies tried to achieve interpretable results, for example, a temporal rule learning methodology was used to extract human-interpretable logical statements for automatic identification of artifacts in monitoring critically ill patients [68]. Fiterau *et al*. showed that the identification of artifacts in real time high frequency vital sign data can be handled automatically and in a human understandable fashion [68]. The performance of the Gaussian Process Regression was investigated for the functional characterisation of vital-sign trajectories [69]. It was shown that the proposed approach was able to discriminate between abnormal patient trajectories corresponding to those who deteriorated physiologically and were admitted to a higher level of care, from those belonging to patients who had no clinically relevant events. Most of the mentioned work can be extended for mortality prediction. And once again, they confirm that big data–driven, ML approach outperforms traditional analytic techniques for predicting in-hospital mortality.

Chapter 4

Preliminary Results: the Analysis of the Data

The focus of this chapter is on the ShockOmics dataset (clinical data) analysis and corresponding preliminary results for outcome prediction. Since the dataset had a lot of missing values, it was important to try out different data imputation techniques and see the results they provide. Here, we compare different multiple imputation techniques, which are divided into two categories: those that work for multiple-types of data (Predictive mean matching, K-Nearest Neighbor, Random Forest, Classification/regression trees) and those that work only with numeric data (Bayesian Linear Regression, Linear Regression ignoring model error, Linear Regression using bootstrap, Linear regression with predicted values). To measure their performance, several machine learning (ML) algorithms were used, with the accuracy and the Matthews correlation coefficient (MCC) as performance measures (their mean $\pm$ st.d.).

The goal of the ML models is to predict the outcome of patients and the feature set for these experiments includes common markers for mortality risk assessment. These features were chosen based on the literature and modern-day practices for shock outcome prediction. After the imputation techniques and ML models were evaluated, the reconstruction errors were calculated to choose the imputation technique to be used in the later experiments. Here, we assume that the imputation techniques that works the best for the chosen subset will be the best for the whole dataset. So, for later experiments only one imputation technique is used.

After reconstruction errors were calculated a few Bayesian Networks with custom structures were tested. The structures were chosen based on casual relationship assumptions, motivated by the pseudo time-series nature of the ShockOmics dataset. The results obtained in these experiments were used as a baseline for the later experiments.

4.1 Dataset Description

As previously mentioned, the ShockOmics dataset was collected within the ShockOmics European research project ("Multiscale approach to the identification of molecular biomarkers in acute heart failure induced by shock", Nr. 602706, http://shockomics.org) in three ICUs: Hopital Erasme, Universite Libre de Bruxelles, Belgium; Hospital Universitari Mutua Terrassa, Spain; and Hopitaux Universitaires de Geneve, Switzerland. A systematic analysis of expression levels of transcripts, genes and their protein products and of peptides generated by proteolysis were carried out on blood samples obtained from ICU patients [10]. The data covered 7 days of hemodynamic signals since the diagnosis of shock (*T0*). Clinical data and blood samples

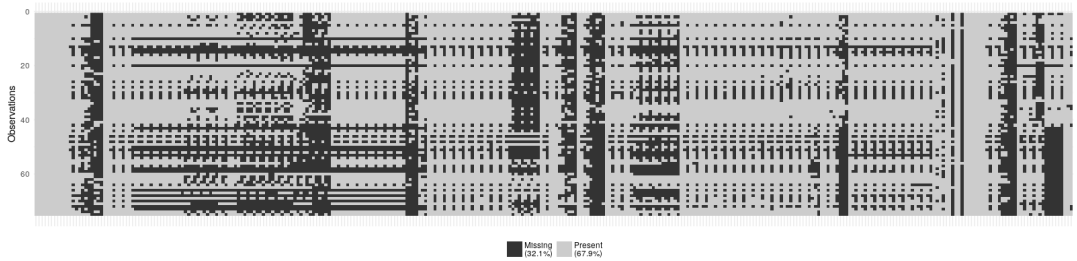


FIGURE 4.1: Observations, features and their missing values in the ShockOmics data. The full list of features and their missing values can be found in Appendix A, Figures A.1, A.2, A.3, A.4

were collected for analysis at: 1) *T1* < 16 h from *T0*; 2) *T2* = 48 h after *T0*; 3) *T3* = day 7 or before discharge or before discontinuation of therapy in case of fatal outcome [10]. Since a lot of patients either died or were discharged during the study, the dataset has a lot of missing values at later time-steps (especially at *T3*).

The ShockOmics data contains scoring systems (e.g. SOFA, APACHE II, Glasgow Coma Scale) and biomarkers on multiple levels. The major focus of the data is on molecular triggers of acute heart failure associated with shock and inflammatory mediators and markers which are activated after an initial insult. According to [70]:

> Several data sources integrate the ShockOmics database: a) plasma samples from existing bio-banks (Albios and Proteosepsis); b) blood samples and hemodynamic recordings from septic shock (SS) and cardiogenic shock (CS) patients, and from sepsis patients; c) bio-samples and hemodynamic data obtained from animal models of shock; d) data from in vitro models of cardiac injury in shock (Neuro-Zone SRL, Milano, Italy).

The clinical data of the ShockOmics dataset, that is used in this thesis (later referred to as the ShockOmics dataset), has 333 features and 75 observations. 32.1% of all values are missing, and the highest percentage of missing values in one feature is 97%. Figure 4.1 shows a graphical depiction map of missing values for each observation (for the full list of features and their corresponding missing values please refer to Appendix A, Figures A.1, A.2, A.3, A.4). The dataset has both numerical and categorical values, but the majority of data is numerical (see Appendix A, Figures A.5).

4.2 Preliminary Experiments Methodology

For a better understanding of the ShockOmics dataset, a series of experiments were conducted. Since the initial dataset had a lot of features and missing values, it was required to choose a subset of features (see Sections 4.2.3 and 4.2.4) and then use imputation techniques (see Section 4.2.1) to get rid off the missing values. Multiple imputation techniques were applied to the data and then evaluated. The evaluation was based on two factors: the first was the performance of the ML models, the second, the reconstruction error.

For the evaluation of the performance of the ML models (see Section 4.2.2) the Accuracy and the MCC were used as performance measures. The training was performed 100 times using 75% of the data. Each time, the split of the data was random. Then the mean and the standard deviation of the 100 executions (using the test set) were calculated for both performance measures.

For the reconstruction error, a different subset of features was used (Section 4.2.4). The chosen subset did not have any missing values. It was artificially reduced to have 10%, 20% and 35% of missing values. Then it was imputed by each of the techniques. Based on the lowest reconstruction error, the imputation techniques were chosen. Additionally, the performance of ML models was similarly measured as in the imputation experiments, for the data with 35% of missing values. Then, these results were compared with the performance of the models using the original dataset with no missing values.

After the data was completed by the imputation technique that showed the best performance, the Bayesian Network structure experiments were conducted. Their purpose was to identify the number of variables discretization and to choose between the Maximum Likelihood Estimator and the Bayesian Estimator. The results of all experiments were designed to evaluate the data and the performance that can be achieved for the outcome prediction task.

4.2.1 Imputation Techniques

Two types of the imputation techniques were used in the experiments. The first one handles multi-type data that can handle missing values of different types (numerical, categorical, etc.). And the second one is only for numerical missing values. In this section, a list of these techniques is presented, including their short description and further notation. The imputations were done in *R* [71] using different packages with imputation techniques. The default parameters of the methods were used.

The experiments include the following imputation techniques for multi-type data and their further notation:

- **pmm**. Imputation by predictive mean matching [72]. A general purpose semi-parametric imputation method. Calculates imputations for univariate missing data by predictive mean matching (*mice* package [73]);
- **knn**. K-Nearest Neighbour imputation based on a variation of the Gower Distance for numerical, categorical, ordered and semi-continous variables (*VIM* package [74]);
- **missForest**. Non-parametric missing value imputation using Random Forest [75]. It can be used to impute continuous and/or categorical data including complex interactions and nonlinear relations (*missForest* package [76]);
- **cart**. Imputation by classification and regression trees [72]. Imputes univariate missing data using classification and regression trees (*mice* package [73]).

The experiments include the following imputation techniques for numeric data and their further notation:

- **norm**. Imputation by Bayesian linear regression [77]. Calculates imputations for univariate missing data by Bayesian linear regression, also known as the normal model (*mice* package [73]);
- **norm.nob**. Imputation by linear regression without parameter uncertainty [73]. Similar to norm, except it does not account for the uncertainty of the model parameters (*mice* package [73]);
- **norm.boot**. Imputation by linear regression using bootstrap [73]. Imputes univariate missing data using linear regression with bootstrap (*mice* package [73]);

- **norm.predict**. Imputation by linear regression through prediction [78]. Imputes the "best value" according to the linear regression model, also known as regression imputation (*mice* package [73]);

4.2.2 Machine Learning Models

The mortality prediction task at hand is a classification problem with two classes: Dead (18 observations) and Alive (57 observations). Several ML algorithms were chosen for the experiments. The *scikit-learn* [79] library was used for their implementation. The following algorithms were used in the experiments:

- **SVC-R**. Support vector machines classification [80] with the radial-basis function kernel, enabled probability estimates and automatically adjusted weights, that are inversely proportional to class frequencies in the input data. The rest of the parameter were used by default. The full name of the method in the *scikit-learn* library is sklearn.svm.SVC (see `http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC`);
- **LDA**. Linear Discriminant Analysis [81]. A classifier with a linear decision boundary, generated by fitting class conditional densities to the data and using Bayes' rule. The model fits a Gaussian density to each class, assuming that all classes share the same covariance matrix. The default parameters were used. The full name of the method in the *scikit-learn* library is klearn.discriminant_-ana-lysis.LinearDiscriminantAnalysis (see `http://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html`);
- **KNN**. Classifier implementing k-nearest neighbors [82]. The number of neighbors was set to three, the rest parameters were set to default. The full name of the method in the *scikit-learn* library is sklearn.neighbors.KNeighborsClassifier (see `http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html`);
- **G-NB**. Gaussian Naive Bayes classifier [83]. The default parameters were used. The full name of the method in the *scikit-learn* library is sklearn.naive_bayes.GaussianNB (see `http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html`);
- **M-NB**. Naive Bayes classifier for multinomial models [84]. The default parameters were used. The name of the method in the *scikit-learn* library is sklearn.naive_bayes.MultinomialNB (see `http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html`);

4.2.3 Imputation Experiments Feature Subset

In the imputation experiments, the subset of features from the ShockOmics dataset was used. As can be seen in Table 4.1, there are 14 features with one target value, namely *Result in ICU*. In order to perform the mortality risk assessment, a few most common biomarkers were chosen. Some of the imputation methods failed to complete *APACHE II* at *T3* and *SOFA* at *T3* columns, so they were left out. Table 4.1 shows the types of the features and the amount of available observations for each column (where 75 means that there is no missing values for the feature). Additionally, it should be noted that the only categorical feature was not included in the experiment with numerical imputation techniques.

Name of the column	Num. of observations	Type
Which type of shock	75	categorical (4 val)
Lactate levels $(mmol/L)_{T1}$	72	numerical (cont)
Lactate levels $(mmol/L)_{T2}$	**64**	numerical (cont)
Lactate levels $(mmol/L)_{T3}$	**38**	numerical (cont)
Mean arterial pressure $(mmHg)_{T1}$	75	numerical (cont)
Mean arterial pressure $(mmHg)_{T2}$	**71**	numerical (cont)
Mean arterial pressure $(mmHg)_{T3}$	**44**	numerical (cont)
$SOFA_{T1}$	75	numerical (cont)
$SOFA_{T2}$	**71**	numerical (cont)
APACHE II_{T1}	75	numerical (cont)
APACHE II_{T2}	**71**	numerical (cont)
Result in ICU	75	categorical (2 val)

TABLE 4.1: Imputation Experiments Feature Subset: 14 feature columns and 1 target column, namely *Result in ICU*. Columns with missing data are highlighted in bold.

Name of the column	Num. of observations	Type
Mean arterial pressure $(mmHg)_{T1}$	49	numerical (cont)
$SOFA_{T1}$	52	numerical (cont)
APACHE II_{T1}	54	numerical (cont)
Result in ICU	75	categorical (2 val)

TABLE 4.2: Reconstruction Error Experiments Feature Subset: 3 feature columns and 1 target column (*Result in ICU*). Each of these features had 75 examples originally.

4.2.4 Reconstruction Experiments Feature Subset

The feature subset for the reconstruction experiments is a reduced version of the subset from the imputation experiments. Only numerical features with no missing values were left (see Table 4.2). Then, from this subset of data and for each column, random observations were replaced by missing values. The percentage of missing values for each feature was different. For the Mean arterial pressure column, 65% of values left, while for the SOFA and the APACHE scores the numbers were 69% and 72%, respectively. This incomplete data was used with different imputation techniques and ML models. The idea was to gauge the performance of the ML models with this artificially reduced dataset.

The reconstruction error was calculated using the same subset of features. Three datasets with different ratios of missing values were obtained by random removal of values from the complete dataset: 35%, 20% and 10% of missing values. For each column, the ratio of missing values was a little bit different but close to the target ratio. After three datasets with missing values were imputed, they were compared with the complete dataset, and the reconstruction error was calculated as the mean squared error.

4.3 Preliminary Experiments

In these experiments, eight imputation techniques were tested (all techniques and their notations are listed in Section 4.2.3). Experiments are organized in two groups.

Models	pmm	knn	missForest	cart
SVC-R	0.7368 ± 0.0	0.7368 ± 0.0	0.7368 ± 0.0	0.7368 ± 0.0
LDA	0.7426 ± 0.0815	**0.8226 ± 0.0664**	0.7805 ± 0.0752	0.7984 ± 0.0791
KNN	0.7647 ± 0.0634	**0.8468 ± 0.0666**	0.8121 ± 0.0719	0.7679 ± 0.0649
G-NB	0.8095 ± 0.0758	0.8258 ± 0.0685	**0.8484 ± 0.0663**	0.8289 ± 0.0742
M-NB	0.7868 ± 0.0833	**0.8126 ± 0.0654**	0.8095 ± 0.079	0.7868 ± 0.083

(A) Here, *G-NB* shows the best result with the *missForest* imputation. However, most of the models show their best performance with the *knn* imputation.

Models	norm	norm.nob	norm.boot	norm.predict
SVC-R	0.7368 ± 0.0	0.7368 ± 0.0	0.7368 ± 0.0	0.7368 ± 0.0
LDA	0.7921 ± 0.0793	0.7889 ± 0.0789	**0.8026 ± 0.0775**	0.8021 ± 0.0774
KNN	0.7847 ± 0.0666	0.7568 ± 0.0701	0.7642 ± 0.068	**0.8432 ± 0.0627**
G-NB	**0.8489 ± 0.0668**	0.8463 ± 0.0769	**0.8489 ± 0.0673**	0.8442 ± 0.0751
M-NB	**0.8205 ± 0.0809**	0.7932 ± 0.0915	0.7942 ± 0.0774	0.8095 ± 0.0705

(B) G-NB shows the best result with the norm.boot and the norm imputations. The norm.predict imputation shows overall improvement in accuracy for all models, but only the KNN model reaches the maximum performance with it.

TABLE 4.3: Imputation experiments: the accuracies (mean ± standard deviation) of the machine learning models for each dataset, imputed with a different technique. For each model the best results are highlighted in bold. Meaning that a certain imputation techniques works the best for the model. The best results in the experiment groups are underlined: (a) for a multi-type imputation group; (b) for a numeric imputation group.

The goal of the first group is to measure the performance of ML models (their description can be found in Section 4.2.2) on the data imputed with different techniques. Here, these experiments are referred to as *imputation experiments*. The second group is called *reconstruction experiments*. It is structured in two parts: the first one is similar to the *imputation experiments* but with the smaller dataset, while the second calculates the reconstruction error of the same dataset imputed with different imputation techniques and for percentages of missing data. The accuracy and the MCC are used to measure the performance of the ML models, while the mean squared error (MSE) is used for the reconstruction error.

4.3.1 Imputation Experiments Results

Once the models had been trained 100 times, using multiple datasets imputed by different methods, the accuracy and the MCC were calculated. The dataset was split, 75% for the training set and 25% for the test set. Each time, the observations were split randomly.

The accuracy measures for the experiments are shown in Table 4.3. The *G-NB* model showed the best accuracy with the norm and the *norm.boot* imputation techniques. The second best accuracy was obtained by the *G-NB* combined with the *missForest* imputation. The *knn* imputation is worth mentioning since it showed very good performance for all ML models and achieved the third best accuracy with the *KNN* model.

Models	pmm	knn	missForest	cart
SVC-R	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
LDA	0.2994 ± 0.2242	**0.5618 ± 0.2022**	0.4335 ± 0.184	0.386 ± 0.2201
KNN	0.2952 ± 0.2165	**0.6059 ± 0.1854**	0.5274 ± 0.2258	0.3011 ± 0.2343
G-NB	0.5741 ± 0.1801	0.5934 ± 0.1983	**0.6471 ± 0.1635**	0.5345 ± 0.2098
M-NB	0.5209 ± 0.1765	0.597 ± 0.1611	**0.5971 ± 0.159**	0.5046 ± 0.1921

(A) The best performance once again was shown by the *G-NB* model with the *missForest* imputation. The same combination showed the best accuracy. Here, the *missForrest* imputation was the best for *G-NB* and *M-NB*, while the *knn* imputation was the best for the *KNN* and the *LDA* models.

Models	norm	norm.nob	norm.boot	norm.predict
SVC-R	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
LDA	0.4467 ± 0.2311	0.4447 ± 0.2056	0.4818 ± 0.1692	**0.486 ± 0.1814**
KNN	0.4293 ± 0.2262	0.2654 ± 0.2455	0.3255 ± 0.1912	**0.5842 ± 0.1875**
G-NB	0.6249 ± 0.1918	0.6267 ± 0.1811	**0.6454 ± 0.1405**	0.6415 ± 0.1645
M-NB	0.6025 ± 0.1723	0.5747 ± 0.1505	0.5711 ± 0.1645	**0.6042 ± 0.1601**

(B) Again the *G-NB* model showed the best results with the *norm.boot* imputation. The same combination that showed the best accuracy. The rest of the models achieved their highest performance with the *norm.predict*.

TABLE 4.4: Imputation experiments: the MCC (mean ± standard deviation) of the machine learning models for each dataset, imputed with a different technique. For each model the best results are highlighted in bold. Meaning that a certain imputation techniques works the best for the model. The best results in the experiment groups are underlined: (a) for a multi-type imputation group; (b) for a numeric imputation group.

Models	Accuracy	MCC
SVC-R	0.6968 ± 0.0428	-0.0933 ± 0.0811
LDA	0.7721 ± 0.0737	0.3434 ± 0.2227
KNN	0.7111 ± 0.0746	0.1924 ± 0.204
G-NB	0.7216 ± 0.1002	0.4401 ± 0.175
M-NB	0.7542 ± 0.0968	0.4695 ± 0.1625

TABLE 4.5: Reconstruction experiments: the accuracies and the MCCs (mean ± standard deviation) of the machine learning models using the original dataset without any missing values. The best accuracy and MCC are underlined.

The MCC measures are shown in Table 4.4. The *G-NB* model outperformed other models, but now the best result was achieved using the *missForest* imputation, followed by the *norm.boot* technique and then the *norm.predict* imputation (with the same *G-NB* model). The MCC shows that the *G-NB* is much better than other models and that the *missForest* and the *norm.predict* imputations performed the best. The *missForest* showed the highest result and with the *norm.predict* all other models aside *G-NB* achieved their best performance.

The imputation experiments results show that the *G-NB* is a clear winner from all other models in terms of the MCC, and it also outperformed other models in terms of the accuracy (with a smaller gap). As for imputation techniques, the *missForest* and the *norm.predict* imputations showed the most promising results.

4.3.2 Reconstruction Experiments: Models Performance

Before calculating the reconstruction error for several artificially made datasets with different percentages of missing data, it was important to see the performance of the ML models on one of such datasets. The dataset with 35% of missing values was chosen for evaluating ML models, and the experiments were done using the same methodology as in the imputation experiments. In the complete dataset a random sample of data (around 35% for each feature) was replaced with missing values. Then this data was imputed by different techniques and evaluated with the same models.

The accuracy and the MCC of the ML models for the original dataset (without missing values) are shown in Table 4.1. The *LDA* performed the best in terms of accuracy and the *M-NB* model showed the highest MCC. Once the performance on the data without missing values had been achieved, the random values were removed for the data, so approximately 35% of missing values was in each column (see Table 4.2).

After removal of some values in the data, the data was imputed by different imputation techniques and the performance was measured again. The accuracy of the models for different imputation techniques can be found in Table 4.6. As can be seen from the table, the *LDA* model and the *norm.predict* imputations worked the best. Additionally, the best performances for most of the ML models were achieved with the *norm.predict* (except the *G-NB* model). The second and the third best accuracies were shown by the *KNN* model and the cart and the *missForest* imputations, respectively.

Models	pmm	knn	missForest	cart
SVC-R	0.7316 ± 0.0629	0.7126 ± 0.0775	**0.7984 ± 0.0552**	0.7932 ± 0.0563
LDA	0.8237 ± 0.0692	**0.8295 ± 0.0679**	0.7995 ± 0.0681	0.8074 ± 0.0658
KNN	0.8074 ± 0.0756	0.8037 ± 0.0818	0.8589 ± 0.0669	**0.8621 ± 0.0677**
G-NB	0.7305 ± 0.1045	0.7384 ± 0.0916	**0.7758 ± 0.0847**	0.7726 ± 0.1104
M-NB	0.7937 ± 0.0854	0.8089 ± 0.0807	**0.8568 ± 0.0691**	0.7879 ± 0.0895

(A) The best accuracy was shown by the *cart* imputation with the *KNN* model. Most of the models performed better with the *missForest* technique.

Models	norm	norm.nob	norm.boot	norm.predict
SVC-R	0.7258 ± 0.0512	0.7258 ± 0.0543	0.7221 ± 0.0269	**0.8195 ± 0.0472**
LDA	0.8279 ± 0.0694	0.8121 ± 0.0679	0.7779 ± 0.0634	**0.8716 ± 0.0629**
KNN	0.8021 ± 0.0748	0.7537 ± 0.0702	0.7332 ± 0.0683	**0.8526 ± 0.0572**
G-NB	**0.8105 ± 0.0809**	0.7426 ± 0.0794	0.7468 ± 0.0873	0.7805 ± 0.0832
M-NB	0.7821 ± 0.0849	0.7637 ± 0.0916	0.7726 ± 0.0905	**0.8447 ± 0.0671**

(B) The majority of models were improved with *norm.predict* imputation. The best result was shown by this technique and the *LDA* model.

TABLE 4.6: Reconstruction experiments: the accuracies (mean ± standard deviation) of the ML models for each dataset, imputed with a different technique. For each model the best results are highlighted in bold. Meaning that a certain imputation techniques works the best for the model. The best results in the experiment groups are underlined: (a) for a multi-type imputation group; (b) for a numeric imputation group.

The MCC measures for these experiments can be found in Table 4.7. Once again the *norm.predict* imputation showed very promising results, having the first, the second and the third best MCCs with the *LDA*, the *M-NB* and the *KNN* models, respectively. Most of the models performed the best with the *norm.predict* imputation.

From the ML models performance comparison it is clearly seen that the *LDA* model and the *norm.predict* imputation stand out. However, the performance of the models with the imputed dataset is higher than with the original dataset. This problem should be addressed by calculating the reconstruction error.

4.3.3 Reconstruction Experiments: Reconstruction Error

The reconstruction error was calculated as the Mean Squared Error (MSE). The elements of the imputed datasets were subtracted from the corresponding elements in the original dataset. Three datasets with the different percentages of missing values were used in the experiments: 35%, 20% and 10%. The reconstruction errors for these datasets are shown in Table 4.8. It is seen that for each of the datasets the best imputation technique is different. For the 35% of missing values the *cart* imputation showed the lowest error, for 20% – the *norm.predict* technique, for 10% – the *knn* imputation.

Since there was no best imputation method in the experiment and since the original ShockOmics data had very different ratios of missing values for different features, it was decided to calculated the sum of the MSEs for all three datasets. The lowest error in all three dataset was achieved by the *norm.predict* imputation, followed by the *missForest* technique. The gap between these two methods is small,

Models	pmm	knn	missForest	cart
SVC-R	0.1218 ± 0.2098	0.113 ± 0.2316	0.3894 ± 0.226	**0.3724 ± 0.2314**
LDA	**0.4908 ± 0.2233**	0.4501 ± 0.1928	0.4583 ± 0.2164	0.4772 ± 0.1771
KNN	0.4704 ± 0.2015	0.4058 ± 0.1786	0.6123 ± 0.2069	**0.6238 ± 0.1713**
G-NB	0.3841 ± 0.1955	0.4102 ± 0.214	**0.4836 ± 0.1923**	0.4736 ± 0.2342
M-NB	0.5209 ± 0.1765	0.597 ± 0.1611	**0.5971 ± 0.159**	0.5046 ± 0.1921

(A) The pair of the *KNN* model and the *cart* imputation showed the best MCC in the multi-type group. However, the *missForest* imputation improved the same amount of models.

Models	norm	norm.nob	norm.boot	norm.predict
SVC-R	0.1165 ± 0.214	0.097 ± 0.2237	-0.0464 ± 0.0681	**0.4854 ± 0.1961**
LDA	0.5427 ± 0.197	0.4309 ± 0.2195	0.3584 ± 0.2213	**0.7115 ± 0.1735**
KNN	0.5115 ± 0.2006	0.3023 ± 0.2036	0.2092 ± 0.2072	**0.6467 ± 0.1671**
G-NB	**0.5426 ± 0.1752**	0.381 ± 0.1959	0.3398 ± 0.1846	0.5075 ± 0.1587
M-NB	0.5 ± 0.191	0.4284 ± 0.1836	0.3916 ± 0.1899	**0.652 ± 0.1684**

(B) Here, the results are almost the same as with the accuracy. The absolute winner is the *norm.predict* imputation with the highest MCC for the *LDA* model.

TABLE 4.7: Reconstruction experiments: the MCC (mean ± standard deviation) of the ML models for each dataset, imputed with a different technique. For each model the best results are highlighted in bold. Meaning that a certain imputation techniques works the best for the model. The best results in the experiment groups are underlined: (a) for a multi-type imputation group; (b) for a numeric imputation group.

and since the *missForest* has an advantage over the *norm.predict* imputation in dealing with multi-type data, a decision was made to use the *missForest* imputation for the later experiments of this thesis.

4.4 Bayesian Network Experiments

After the data was imputed with the norm.predict imputation (all features with missing values were numerical), different Bayesian Networks Structures were tested for the outcome prediction. The structures of the chosen Bayesian Networks can be seen in Figure 4.2. Each of the configurations were given a number for a later use in the table with the results. The Bayesian Networks had different features and the independence assumptions, that were based on time-series nature of the ShockOmics dataset. Additionally, it was important to see the number of variable discretization (the maximum amount of different values that can be assigned to one variable). For this role the range from 2 to 5 was tested. Additionally, two different estimators were used: the Maximum Likelihood Estimator and the Bayesian Estimator. The accuracy, the MCC, the sensitivity, the specificity and the AUC were used as performance measures. All the models were trained ten times on randomly split dataset, similarly as was done in the previous experiments.

4.4.1 Results

Table 4.9 presents the results of applying the Maximum Likelihood and the Bayesian estimators. The experiments show that the structure 1 was the best among all other

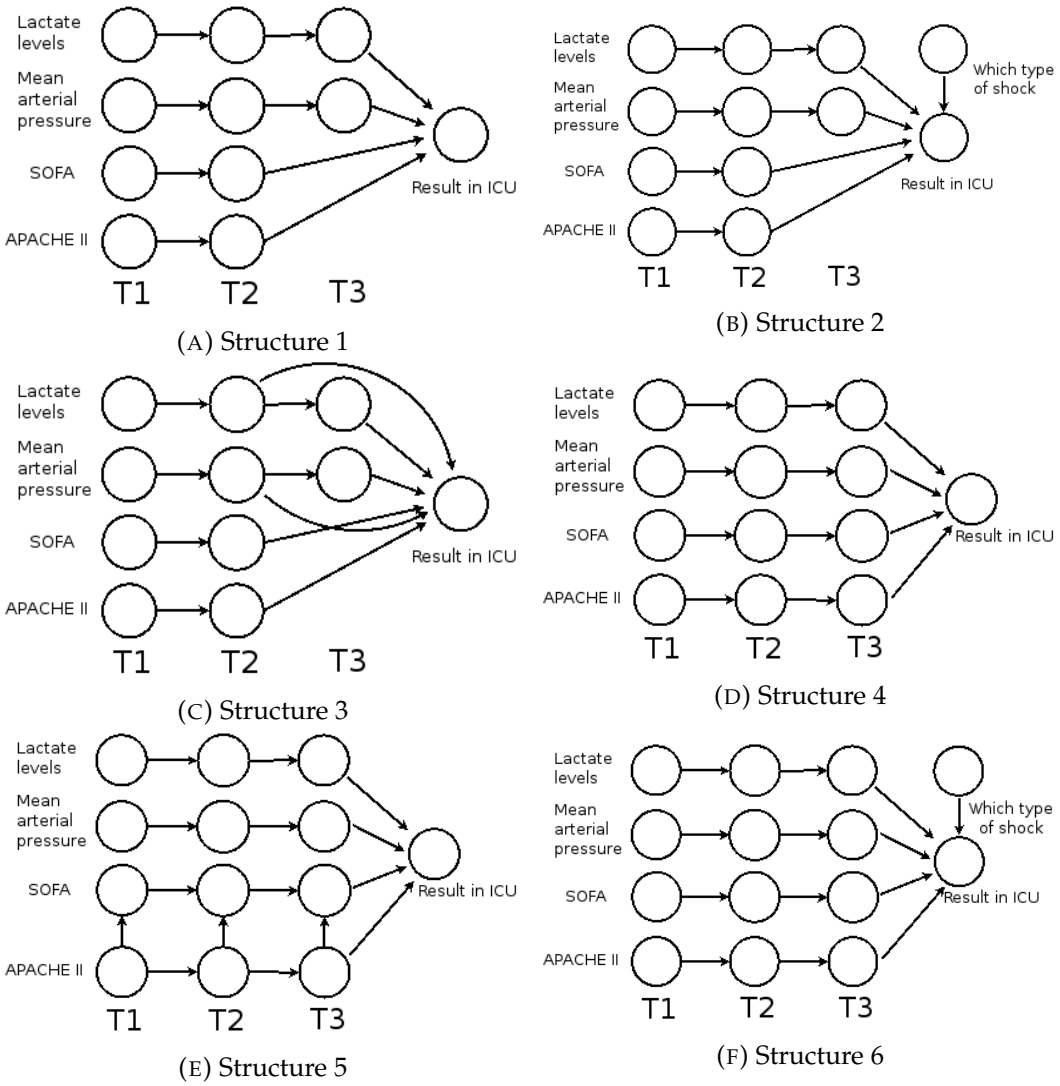
	35% missing	20% missing	10% missing	sum
Num.of examples	49, 52, 54	61, 59, 57	68, 65, 67	
pmm	8671	5083	3226	16980
knn	6556	2784	**838**	10178
missForest	4753.7	3842.9	1073.5	9670.1
cart	**4524**	6527	866	11917
norm	7831.4	5097.9	3184.8	16114.1
norm.nob	6482.8	4371.5	1866.9	12721.2
norm.boot	12124.3	5706.7	1795.9	19626.9
norm.predict	5917.42	**2437.7**	1005.9	**9361.02**

TABLE 4.8: Reconstruction experiments: the MSE for datasets with missing values and the imputation techniques. *Num. of examples* shows the number of examples for each of the features in the following order: *Mean arterial pressure* at *T1*, *SOFA* at *T1*, *APACHE II* at *T1*. The *sum* column represents the sum of all three MSEs for each imputation technique. The best results for each column are highlighted in bold.

configurations. It showed the best accuracy, the MCC and the AUC, as well as the best specificity (see Tables A.2 and A.4 from Appendix A). Three and four discretization variables scored the best performance for each of the six structures. Based on the overall performance, the structure 1 had the biggest success with three discretization variables and the Bayesian Estimator. It showed the best sensitivity/specificity ratio in the experiments.

The table shows that, in general, the Bayesian Estimator performed better than the Maximum Likelihood Estimator. And from these results one can conclude that adding $APACHE\ II_{T3}$ and $SOFA_{T3}$ features (in structures 4, 5 and 6) made the performance of the Bayesian Networks worse. Moreover, one may notice that all models in the experiments have high standard deviations. This inconsistency can be explained by the fact that the ShockOmics dataset has a low number of observations. There were much more positive examples compared to the negative ones in the experiments, so the models tended to have a high sensitivity and low specificity.

As already mentioned, the best results were yielded by the Bayesian Network with three discretization variables and the Bayesian Estimator. These results are comparable with other ML methods that were tested previously in the imputation experiments. The accuracy of the best ML model in those experiments (*G-NB*) is comparable of the results achieved by the best Bayesian Network performance. It is 0.8632 ± 0.0573 here, against 0.8489 ± 0.0673 in the *G-NB* model. However, the Bayesian Network shows the lower MCC: 0.6181 ± 0.0854 against 0.6471 ± 0.1635. These experiments show that the Bayesian Networks are not only capable of shedding some light on casual relationship between the features, but can also perform in pair to other ML models even with very general independence assumptions.



(A) Structure 1

(B) Structure 2

(C) Structure 3

(D) Structure 4

(E) Structure 5

(F) Structure 6

FIGURE 4.2: Bayesian Network structured from the Bayesian Network experiments. Each horizontally connected row represents the same feature at different time steps. All features at the last time step are connected to the outcome.

Struct.	Discr.	Accuracy	MCC	AUC
1	3	**0.8526 ± 0.0614**	0.6023 ± 0.1441	**0.8526 ± 0.0614**
1	4	**0.8526 ± 0.1219**	**0.6381 ± 0.2973**	**0.8526 ± 0.1219**
2	3	0.8421 ± 0.0471	0.5723 ± 0.1157	0.8421 ± 0.0471
2	4	0.8 ± 0.0614	0.1063 ± 0.1399	0.8 ± 0.0614
3	3	0.7895 ± 0.0333	0.364 ± 0.1988	0.7895 ± 0.0333
3	4	0.8421 ± 0.0881	0.2749 ± 0.3367	0.8421 ± 0.0881
4	3	0.7053 ± 0.0537	0.0585 ± 0.1893	0.7053 ± 0.0537
4	4	0.7368 ± 0.0881	0.1405 ± 0.2938	0.7368 ± 0.0881
5	3	0.6316 ± 0.0577	0.0337 ± 0.1509	0.6316 ± 0.0577
5	4	0.7053 ± 0.0976	0.0009 ± 0.1749	0.7053 ± 0.0976
6	3	0.7263 ± 0.0394	0.0094 ± 0.1194	0.7263 ± 0.0394
6	4	0.7474 ± 0.0394	-0.0092 ± 0.1216	0.7474 ± 0.0394

(A) Most of the structures work the best with three discretization variables. The first three structures performed better than the last three. Full Maximum Likelihood Estimator results can be found in Tables A.1 and A.2.

Struct.	Discr.	Accuracy	MCC	AUC
1	3	**0.8632 ± 0.0537**	**0.6181 ± 0.0854**	**0.8632 ± 0.0537**
1	4	0.8421 ± 0.0577	0.5863 ± 0.1127	0.8421 ± 0.0577
2	3	0.8 ± 0.0774	0.4197 ± 0.1773	0.8 ± 0.0774
2	4	0.8421 ± 0.0744	0.368 ± 0.2782	0.8421 ± 0.0744
3	3	0.8526 ± 0.0211	0.1882 ± 0.2306	0.8526 ± 0.0211
3	4	0.7158 ± 0.0855	0.153 ± 0.1931	0.7158 ± 0.0855
4	3	0.7053 ± 0.0632	0.0924 ± 0.168	0.7053 ± 0.0632
4	4	0.7368 ± 0.0577	0.0502 ± 0.2463	0.7368 ± 0.0577
5	3	0.7158 ± 0.0421	0.0056 ± 0.1152	0.7158 ± 0.0421
5	4	0.7579 ± 0.0537	0.1688 ± 0.2054	0.7579 ± 0.0537
6	3	0.7263 ± 0.0774	0.11 ± 0.1692	0.7263 ± 0.0774
6	4	0.7474 ± 0.0842	0.2249 ± 0.1893	0.7474 ± 0.0842

(B) The results are very similar to those shown by the Maximum Likelihood Estimator. The gap in the performance between the second and the third structures is more clearly here (the structure 2 performed better). Full Bayesian Estimator results can be found in Tables A.3 and A.4.

TABLE 4.9: Different Bayesian Networks structures with different numbers of discretization variables: a) for the Maximum Likelihood Estimator; b) for the Bayesian Estimator. The best results in each column are highlighted in bold. Full results for the experiments (with additional specificity and sensitivity measures) are in Appendix A.

Chapter 5

Feature selection: Identifying Promising Biomarkers

This chapter is dedicated to the data feature selection procedure from an ML point of view. Therefore, it deals with the problem of finding the most promising biomarkers from the medical point of view. In order to do this, four subsets of the ShockOmics data were analyzed, giving them different names: **Full+**, **Full**, **T1+T2** and **T1** (description in the following sections). The goal of the experiments was to analyze different feature sets resulting from the feature selection procedure, and to measure their performance with the Gaussian Naive Bayes (*G-NB*) model. There is an assumption behind the experiments, namely that the feature sets with the best performances are the same those that can potentially help with the mortality prediction.

The evaluated feature sets were obtained by applying five different feature selection techniques to the four mentioned datasets: Univariate feature selection based on ANOVA F-value (*UFS*); recursive feature elimination with Support Vector Classification (*RFE*); recursive feature elimination with cross-validation and Support Vector Classification (*RFECV*), two stage feature selection (*UFS+RFE*) and Random Forest (*RF*) feature selection. The dataset was imputed with the *missForest* imputation, and all the ShockOmics features were used if not specified otherwise. The identification of the promising feature sets was based on the performance and stability scores.

The chapter is organized as follows. The first section contains all the details about the experimental methodology. It starts with a description of all the datasets and then describes the feature selection methods used, as well as how the resulting feature sets were evaluated and how the stability scores were calculated. The next section discusses the results of applying the feature selection procedure for each of the datasets in particular and presents their stability scores. After this, the final section presents the final sets of promising features for further experiments.

5.1 Feature Selection Methodology

In order to identify promising feature sets, five feature selection techniques (see Section 5.1.2) were applied to four subsets of the ShockOmics datasest (see Section 5.1.1). The resulted feature sets were used to train the *G-NB* model from the previous chapter. Then, the models were evaluated as in the previous chapter (see Section 5.1.3). The accuracy, the MCC, the specificity, the sensitivity and the AUC were used as performance measures. The training was performed 100 times using 75% of the data, that was randomly split. Then the mean and the standard deviation of the 100 executions (using the test set) were calculated for all five performance measures.

Once all feature sets, obtained with the feature selection techniques, were evaluated, they were used to calculate the stability scores (see Section 5.1.4). The stability

scores were based on the occurrences of features throughout the feature selection results. The stability scores for the first four features and for the *RF* feature selection were calculated separately, since the *RF* is the only stochastic technique in the experiments and was treated differently.

5.1.1 Datasets Description

The decision of using four different subsets of the ShockOmics dataset was made because it was important to achieve stable features that behave consistently at different time-steps and that are clinically relevant. For this reason, each of the datasets targeted different features: some of them because of their low amount of missing values, some of them because they were important at different time. The datasets were obtained by imputing missing values from the ShockOmics dataset with the *missForest* imputation from the previous chapter. All the datasets contain both numerical and categorical features, and they are subsets of the ShockOmics dataset.

The first dataset was named **Full+**. It was achieved by filtering features from the ShockOmics dataset (that had 333 features). It was obtained by the manual removal of certain features that did not make sense from the classification point of view. They either looked like very general comments in natural language or revealed some information about the outcome of the patient. Both of the categories were not reliable as features in the machine learning model. The following features were manually removed from the ShockOmics dataset: *Reason for admission*; *ICU admission*; *RV area/LV area* (*T1, T2, T3*); *Microorganisms* (three columns with the same name); *ID*; *Death due to withdrawal of care*; *Mortality 28 days, 100 days* (two columns); *Hospital results*; *Total days in ICU, in Hospital* (two columns). The resulting dataset had 317 features with one target feature.

The second dataset is the **Full** dataset. Its purpose was to address certain drawbacks of the **Full+** dataset by removing the features with a lot of missing values. From all the list of the **Full+** features, only those that had more than 50% values originally were taken. The imputation techniques are not accurate when there are more missing than present values, especially when it comes to datasets where the amount of observation is very low. So, the removal of the features with too many missing values was used to increase the stability of the promising features in the end of the experiments. After filtering, the dataset **co**ntained 242 features with one target feature.

Additionally, there was an assumption that the closer in time the feature to the final outcome, the better the prediction of the mortality. This assumption was addressed by making two last datasets: the **T1+T2** dataset and the **T1** dataset. both of them were obtained from the **Full+** dataset by filtering features at certain time-steps. From the **T1+T2** dataset all T3 features were removed, resulting in 180 features with a target feature. From the **T1** dataset both T2 and T3 time-step features were left out, and the resulted dataset contained 104 features and a target feature.

5.1.2 Feature Selection Methods

The scikit-learn [79] library was used for the implementation of the feature selection methods. The default parameters of the methods were used in the experiments, if not otherwise specified. Five feature selection techniques were used in the identification of promising biomarkers:

- **Univariate feature selection based on ANOVA F-values (UFS).** Univariate feature selection [85] works by selecting the best features based on univariate

statistical tests. In this case, the ANOVA F-tests are used. The *UFS* was implemented using the sklearn.feature_selection.SelectKBest method that removes all but the k highest scoring features (see http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html#sklearn.feature_selection.SelectKBest) and sklearn.feature_selection.f_classif for the ANOVA F-tests (see http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_classif.html#sklearn.feature_selection.f_classif) ;

- **Recursive feature elimination with SVC (RFE).** Given an external estimator that assigns weights to features (e.g., the coefficients of a linear model), the goal of the *RFE* is to select features by recursively considering smaller and smaller sets of features [86]. The *RFE* was implemented through the sklearn.feature_-selection.RFE method (http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html). For the external estimator, the SVC model [80] was used with the help of the sklearn.svm.SVC method (see http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html);

- **Recursive feature elimination with cross-validation and SVC (RFECV).** Feature ranking with the *RFE* and cross-validated selection of the best number of features. Different scores were used of the cross-validation: AUC, recall, accuracy, precision, average precision, *neg log loss*, *f1*, *f1 micro*, *f1 macro* (the description of these scores can be found in http://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter. The sklearn.-feature_-selection.RFECV method allowed to use this feature selection technique (see http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFECV.html). Once again, the SVC model was used as the external estimator (see previous entry);

- **Two stage feature selection: UFS + RFE.** This is a combination of two feature selection techniques from this list. During the first stage, UFS preselects the 80 best features based on the ANOVA F-value and then *RFE* (SVC) choses the best features in the second stage. For the implementation details, please, refer to the individual techniques from above this list;

- **Random Forest (RF) feature selection.** Feature selection [86] based on the trained *RF* model [62] and its feature rankings. In the experiments, this feature selection was based on 500 runs of the *RF* model with 10 trees. This was carried out because the *RF* feature selection was the only stochastic technique, and multiple experiments were required to achieve a high level of certainty in the results. The final feature selection was based on the stability scores from all the executions. The most frequent features got in the final feature subset and all the performance evaluation was done using this feature set. The sklearn.feature_selection.SelectFromModel method was used to retrieve the feature rankings (see http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectFromModel.html) and the sklearn.ensemble.RandomForest-Classifier (see http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html) method to train the *RF* model.

5.1.3 Feature Sets Evaluation

As an output, the applied feature selection techniques present either an unranked (e.g. *RFECV*), or a ranked (the rest of the methods) set of features. In the first case, there is a set of features that is fixed in size. Based only on the output, there is no way of identifying which features are more important. And from the feature selection point of view, these features play equally important roles in the set. For the second case, it is different and all features get their own rank. So, naturally, and in order to limit the size of the selected features, only the first ones should be selected and evaluated. But the number of features that should be used for the training is unclear. That is why for the feature selection techniques that provide the rank of features, multiple feature set sizes in the range from 2 to 30 were tested. The upper boundary 30 was chosen because the final set of features should be relatively small to be usable and effective enough in the ML model.

Each feature set was used for the training of the *G-NB* model and carefully evaluated. The *G-NB* classifier was chosen because it showed promising results in the preliminary experiments and because technically it can be considered as a special case of Bayesian Networks. Unfortunately, conducting all the experiments with Bayesian Networks was unfeasible because there was no clear understanding of the independence assumptions. Therefore, the *G-NB* model was used as the substitution. In order to evaluate the feature sets, the *G-NB* was trained and tested 100 times for each one of them. The mean and the standard deviation from all the executions were subsequently calculated. The results are presented separately for each dataset and each feature selection method. The accuracy, the MCC, the specificity, the sensitivity and the AUC were used as performance measures.

5.1.4 Stability Scores

Once all the feature sets were tested and their performances were measured, the stability scores were calculated. For each of the dataset there were two groups of stability scores. The first group contained only features obtained with the *RF* feature selection. The second group had the rest of the methods. The reason behind the difference between these two groups is that the *RF* feature selection is stochastic. Therefore, it was executed 500 times and each time it gave a slightly different feature set. The methods in other group were applied only once, since the result is always the same. By separating two groups, and by giving their stability scores the equal weight in the decision making, the difference of the methods was compensated.

In each group, the stability scores were calculated as frequencies. The number of individual feature occurrences was divided by the number of feature selection procedure executions in the group. So, in the *RF* group it was divided by 500, and in the second group it was divided by three plus the number of the *RFECV* feature sets. The number of the *RFECV* feature sets varied depending on the dataset. This is because different scores may lead to different feature sets. And only the number of unique feature sets was used for calculating the stability scores. In the end each dataset had two groups of stability scores. And each group was represented using 30 features with the highest stability scores.

5.2 Feature Selection Results

The feature selection experiments are divided into four groups. Each group corresponds to its own dataset: **Full+**, **Full**, **T1+T2** and **T1**. All five feature selection techniques were applied to the datasets. The results were used to train the *G-NB* model, and the performance of each feature set with the model was evaluated. In this section, the best models from all feature selection experiments and their accuracy, MCC and AUC are presented. For the complete experiments, results and additional measures as the sensitivity and the specificity, please refer to the individual subsections and corresponding appendix sections. They include full results for each dataset.

Table 5.1 shows the comparison of different feature sets: their size, the dataset and the feature selection method that was used. The feature sets from the same dataset are grouped together. The additional IFS dataset in the table corresponds to the feature set from the preliminary experiments, it contains 13 features and it is used as a baseline for other models performance. It was interesting to compare the usual common features for mortality prediction and the new feature sets. The table presents the best results for each feature selection method.

As can be seen from Table 5.1, the *UFS* and the *RF* models show consistently good results. Their performance is better than the models with the IFS feature set, and this is especially noticeable when it comes to the MCC. The best results in the experiments were found for the *RF* feature selection in the **T1+T2** dataset. This is interesting, and it may be interpreted as the T3 time-step being less significant for the outcome prediction. An alternative interpretation is that there is too much noise in the T3 features, since the patient with shock becomes very unstable at that point, or maybe because there are too few observations at that time-step. The **T1+T2** feature sets performed very well despite the absence of the T3 time-step.

The *RFECV* features showed the worst results. This is partially because the SCV model was used for cross-validation, but the performance was measured with the *G-NB*. It was impossible to use the *G-NB* model, since it can not evaluate the importance of features. When the SVC model was used for testing the performance, it showed slightly better results. As for the *USF+RFE* selection, it performed slightly worse than the UFS in general. The *RFE* feature showed reasonably good performance but worse than the UFS+RFE.

5.2.1 Full+ Dataset Results

This subsection describes the experimental results for the **Full+** dataset. The complete set of tables from this subsection can be found in Appendix B. Table B.1 shows the performance of the models that used the *UFS* technique for feature selection. The models with 9 and 10 features performed the best. The first one has the best MCC and the sensitivity and the second one outperformed all other models in the accuracy and the AUC. The features in these models can be found in Table B.6, the first 9 and 10 features respectively.

The performance of models that used the *RFE* feature selection can be seen in Table B.2. Here, two models performed well compared to others. The first model has 16 features and it showed the best accuracy and the AUC. The second model has 20 features and it has the best MCC. The features of these models can be found in Table B.6. As for the *RFECV* features, their performance is in Table B.3. There are only two models since this method uses cross-validation to choose the best set of features. The first model used the *average precision* metric, the second used the

F	Data	Method	Accuracy	MCC	AUC
13	IFS	-	0.802 ± 0.081	0.403 ± 0.327	0.802 ± 0.081
9	Full+	UFS	0.862 ± 0.069	0.667 ± 0.168	0.862 ± 0.069
10	Full+	UFS	0.863 ± 0.078	0.663 ± 0.194	0.863 ± 0.078
16	Full+	RFE	0.765 ± 0.083	0.359 ± 0.237	0.765 ± 0.083
20	Full+	RFE	0.763 ± 0.076	0.354 ± 0.21	0.763 ± 0.076
17	Full+	RFECV	0.748 ± 0.081	0.31 ± 0.233	0.748 ± 0.081
19	Full+	RFECV	0.719 ± 0.084	0.228 ± 0.235	0.719 ± 0.084
4	Full+	UFS+RFE	0.842 ± 0.104	0.6 ± 0.235	0.842 ± 0.104
30	Full+	UFS+RFE	0.841 ± 0.079	0.605 ± 0.189	0.841 ± 0.079
28	Full+	RF	0.861 ± 0.062	0.655 ± 0.161	0.861 ± 0.062
24	Full	UFS	0.872 ± 0.073	0.687 ± 0.173	0.872 ± 0.073
11	Full	RFE	0.822 ± 0.088	0.559 ± 0.222	0.822 ± 0.088
13	Full	RFE	0.828 ± 0.083	0.546 ± 0.229	0.828 ± 0.083
12	Full	RFECV	0.788 ± 0.088	0.468 ± 0.223	0.788 ± 0.088
23	Full	RFECV	0.756 ± 0.071	0.292 ± 0.247	0.756 ± 0.071
25	Full	UFS+RFE	0.803 ± 0.071	0.395 ± 0.27	0.803 ± 0.071
30	Full	UFS+RFE	0.799 ± 0.06	0.421 ± 0.221	0.799 ± 0.06
7	Full	RF	0.866 ± 0.075	0.68 ± 0.179	0.866 ± 0.075
15	Full	RF	0.866 ± 0.075	0.673 ± 0.18	0.866 ± 0.075
16	Full	RF	0.866 ± 0.07	0.663 ± 0.169	0.866 ± 0.07
28	T1+T2	UFS	0.855 ± 0.06	0.64 ± 0.155	0.855 ± 0.06
27	T1+T2	RFE	0.819 ± 0.085	0.561 ± 0.212	0.819 ± 0.085
41	T1+T2	RFECV	0.812 ± 0.078	0.534 ± 0.201	0.812 ± 0.078
28	T1+T2	UFS+RFE	0.791 ± 0.052	0.371 ± 0.216	0.791 ± 0.052
5	T1+T2	RF	**0.874 ± 0.073**	**0.693 ± 0.181**	**0.874 ± 0.073**
30	T1	UFS	0.867 ± 0.063	0.668 ± 0.158	0.867 ± 0.063
20	T1	RFE	0.764 ± 0.067	0.197 ± 0.296	0.764 ± 0.067
29	T1	RFE	0.755 ± 0.058	0.207 ± 0.253	0.755 ± 0.058
31	T1	RFECV	0.75 ± 0.069	0.203 ± 0.252	0.75 ± 0.069
21	T1	UFS+RFE	0.791 ± 0.043	0.357 ± 0.206	0.791 ± 0.043
29	T1	UFS+RFE	0.776 ± 0.072	0.393 ± 0.198	0.776 ± 0.072
4	T1	RF	0.848 ± 0.078	0.621 ± 0.206	0.848 ± 0.078

TABLE 5.1: The best feature sets of the feature selection techniques, and their comparison to the initial feature set (IFS): the size of the feature set, the dataset, the feature selection technique and different performance measures. For full experiment results, please, see Appendices B, C, D, E.

accuracy. It was important to test both models, since they showed very different set of features. These sets can be found in Listing B.2.

Finally, Table B.4 shows the two-stage feature selection results. Two feature sets with different sizes of 4 and 30 stand out. Table B.5 presents the performance of the features chosen using the RF. The 28 feature sets showed the best performance foR this feature selection technique. The list of features for both techniques can be found in Table B.7.

Table F.1 from Appendix F shows the stability scores for the **Full+** dataset.

5.2.2 Full Dataset Results

Appendix C contains all the tables that are discussed here. Table C.1 shows the performance of the models that used the *UFS* technique. The model with 24 features performed the best. The features that belong to this model can be seen in Table C.6, the first 24 features. The performance of models that used the *RFE* feature selection can be seen in Table C.2. Two models are considered to be the best ones here. The first model has 11 features and it showed the best MCC. The second model has 13 features and it has the best accuracy and the AUC. The features of these models can be found in Table C.6, the first 11 features for the first model, 13 – for the second.

The results of the experiments using the *RFECV* are to be found in Table C.3. There are only two models, since this method uses cross-validation to choose the best set of features. The first model used the *precision* metric, the second used the *AUC*. Both feature sets can be found in ListingC.2. Table C.4 shows the two-stage feature selection results. Table C.5 presents the performance of the features chosen using the RF. The list of features for both models can be found in Table C.7.

Table F.2 from Appendix F shows the stability scores for the **Full** dataset.

5.2.3 T1+T2 Dataset Results

The **T1+T2** dataset result tables and listing are to be found in Appendix D. Table D.1 shows the results for the *UFS* feature selection. It seems that the best results were produced by feature sets with relatively high sizes. For example, the feature set with 28 showed the best accuracy and AUC. And the size 26 feature set achieved the best MCC which is also a very important performance measure. The rankings of features selected by the *UFS* are shown in Table D.6.

The *RFE* performance is shown in Table D.2. As can be seen, the situation is similar here: the feature sets with high sizes score better. The best results were achieved with the size 27. This feature set showed the best accuracy, MCC and AUC. The sizes 26 and higher showed very similar results. The ranking of the features selected by the *RFE* can be found in Table D.6.

The *RFECV* extracted 5 different feature sets using multiple metrics. Table D.3 shows the performance of these sets. The best candidate here for a feature set is the set with the size 41. It shows the best accuracy, MCC, sensitivity and AUC. However, these results are much lower than those shown by the *UFS* and the RFE. These feature sets are in Listing D.2.

The *USF+RFE* showed worse results than the *UFS*, but the same amount of features showed the best performance. The feature set with the size 28 achieved the best accuracy, MCC, sensitivity and AUC. The ranking of features achieved with this method are shown in Table D.7 along with the *RF* features. The *RF* features results can be found in Table D.5. A feature set with only five features achieved significant performance for such small set.

Table F.3 from Appendix F shows the stability scores for the **T1+T2** dataset.

5.2.4 T1 Dataset Results

The **T1** dataset results can be found in tables and listing from Appendix E. This time the *UFS* performed the best among other methods. However, it took 30 features to outperform other candidates. Table E.1 shows the results of applying this method to the **T1** dataset. As can be seen, the feature set with the size 30 performed the best. It has the highest accuracy, the MCC and the AUC. The ranking of features obtained by this method are in Table E.6.

The *RFE* with **T1** performed quite poorly in comparison to other datasets. Table E.2 shows the performance of the chosen feature sets. From the reported results, it seems that there is no clear best performance. The MCC and sensitivity is very low among all feature sets. The feature set with the size 20 showed the highest accuracy, the AUC and the second best MCC, so it may be considered as the best feature set shown in the table. The ranking of features for this method can be found in Table E.6.

This time, the *RFECV* showed 6 different feature sets that were achieved with multiple metrics. The performance of these feature sets can be found in Table E.3. The *average precision* metric was used for the best feature set that scored the best accuracy, MCC and AUC. The feature sets for all metrics can be found in Listing E.2.

The *USF+RFE* feature sets performed slightly better than feature sets from the last two methods. From Table E.4 one may notice that all feature sets performed very similarly. It is difficult to point the best feature sets but feature sets with sizes 21 and 29 could be the candidates. Since the first one showed the highest accuracy and AUC while the second one performed the best in the MCC measure. The ranking of features achieved with this methods can be found in Table E.7. The performance of the *RF* features is in Table E.5 with the features ranking in Table E.7. This time the performance was lower than in the previous subsection, however, the size was smaller too.

Table F.4 from Appendix F shows the stability scores for the **T1** dataset.

5.3 Promising Features Sets

In order to obtain the most promising feature sets, the best results from the feature selection experiments were filtered based on their size and performance. During this procedure, a requirement was to prioritize small feature sets (with less than 20 features). This is because using big feature sets is not practical for outcome prediction: it is more difficult to train a model with a high number of features and it is harder to deal with independence assumptions for the Bayesian Network. Furthermore, the resulting models might be more difficult to interpret. As for the performance, the combined filtering was used. Only the feature sets with minimum accuracy ≥ 0.85, MCC ≥ 0.6, sensitivity and specificity ≥ 0.75 and AUC ≥ 0.85 were considered as promising ones. Because, for example, if the accuracy is high and the specificity is very low, then this is probably a bad model, since there are too few negative examples in the ShockOmics dataset.

The performance of the filtered feature sets can be found in Table5.2 along with the best feature sets performance from the preliminary experiments (IFS). The feature sets themselves can be found in Listing 5.3. As can be seen, new feature sets outperform the IFS features. The lists of new feature sets are interesting. The usual

indicators as *SOFA* or *APACHE II* are present in the feature sets. And the new features (e.g. *X Norepinephrine*) have fairly high stability scores. Interestingly, the *RF* stability scores turned out to be much better at predicting valuable features. The features in the filtered sets have not only high performance, but they also go along with the traditional practices for outcome prediction and stability scores (especially, for the *RF* feature selection).

The promising feature sets identified in this chapter are used for exploring their causal relationships in the Bayesian Network in the next chapter.

```
Full+ (9 and 10, UFS): APACHE II_T1, APACHE II_T2, APACHE II_T3,
    Inferior vena cava distensibility index _T3, LA dilatation by
    eyeballing_T1=No, LA dilatation by eyeballing_T1=Yes, SOFA_T1,
    SOFA_T2, SOFA_T3, X Norepinephrine mg kg min _T2;

Full (24, UFS):APACHE II_T1, APACHE II_T2, APACHE II_T3, FiO2_T2,
    FiO2_T3, Glasgow Coma Scale_T2, Glasgow Coma Scale_T3, Heart rate
    bpm _T3 1, LA dilatation by eyeballing_T1=No, LA dilatation by
    eyeballing_T1=Yes, Lactate levels mmol L _T2, Number of affected
    organs, Respiratory rate_T1, SOFA_T1, SOFA_T2, SOFA_T3, Sat O2
    FiO2_T2, Sat O2 FiO2_T3, Sedation Scale SAS _T2, Tidal volume VT
    _T1, Tracheal Intubation_T2=No, Tracheal Intubation_T2=Yes, X
    Norepinephrine mg kg min _T1, pH_T2;

Full (7, 15, RF): SOFA_T2, APACHE II_T1, APACHE II_T3, Lactate levels
    mmol L _T2, Urine Output mL day _T1, Heart rate bpm _T3,
    Respiratory rate rpm _T1, APACHE II_T2, Urine Output mL day _T2, X
    Norepinephrine mg kg min _T1, E wave cm s _T1, FiO2_T3, Sat O2
    FiO2_T3, SOFA_T1, SOFA_T3;

T1+T2 (5, RF): Lactate levels mmol L _T2, SOFA_T2, Respiratory rate rpm
     _T1, APACHE II_T1, E wave cm s _T1.
```

List of promising feature sets. They are presented as the dataset, the size of the feature set (if more than one size is in the parenthesis then there are two different feature sets with two different sizes) and the feature selection technique that was used.

.

F	Data	Method	Accuracy	MCC	AUC
13	IFS	-	0.802 ± 0.081	0.403 ± 0.327	0.802 ± 0.081
9	Full+	UFS	0.862 ± 0.069	0.667 ± 0.168	0.862 ± 0.069
10	Full+	UFS	0.863 ± 0.078	0.663 ± 0.194	0.863 ± 0.078
24	Full	UFS	0.872 ± 0.073	0.687 ± 0.173	0.872 ± 0.073
7	Full	RF	0.866 ± 0.075	0.68 ± 0.179	0.866 ± 0.075
15	Full	RF	0.866 ± 0.075	0.673 ± 0.18	0.866 ± 0.075
5	T1+T2	RF	$\mathbf{0.874 \pm 0.073}$	$\mathbf{0.693 \pm 0.181}$	$\mathbf{0.874 \pm 0.073}$

TABLE 5.2: Feature sets from Table 5.1, filtered according to the performance measures: the accuracy ≥ 0.85, the MCC ≥ 0.6, the sensitivity and the specificity ≥ 0.75 and the AUC ≥ 0.85.

Chapter 6

Causal Discovery: The Bayesian Network Approach

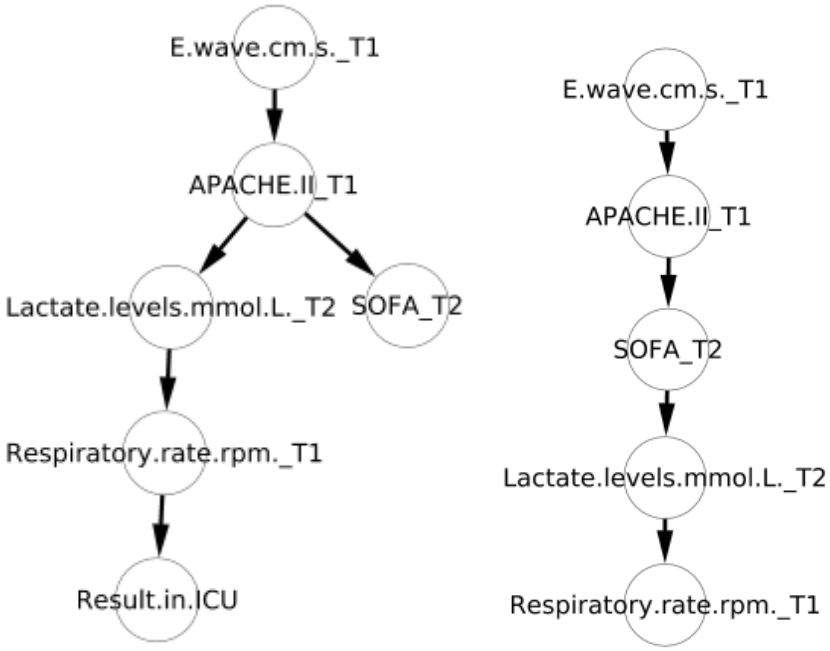
This chapter presents the results of the analysis of potential causal relationships between the data features that were obtained in the previous chapter. Having a set of potential biomarkers is not enough for mortality prediction. It is important to know the causal relationships between them. Different structure learning algorithms for Bayesian Networks can potentially reveal this information. Not only Bayesian Networks allow to see how the biomarkers influence each other and the outcome, but they also work with the partial information. And in the medical environment it is often very difficult to obtain the full picture of the situation and have all feature values.

The feasibility and interest of the use of Bayesian Network models in the medical field has been tested in numerous studies. Some examples amongst the most recent studies include: causal relationships in patients with primary OCD and co-morbid depression symptoms [4]; biophysical interactions of radiation pneumonitis in non-small-cell lung cancer [5]; and identifying causative genome alterations within individual tumors [6]. Given the high effectiveness of the Bayesian Network approach, it was decided to use it for identifying the interactions between the biomarkers that were selected in the experiments reported in the previous chapter.

This chapter includes a single section containing all the details of the casual discovery experiments. A description of the methodology is provided first. This is followed by a short overview of the Fast Greedy Search (FGES) algorithm for structure learning. The second and last subsection presents the structures that were achieved with this algorithm.

6.1 Causal Discovery Experiments

In order to achieve the causal Bayesian Network (CBN) structures that would reveal the interactions between the features, multiple algorithms were applied. Some of them, like the Max-Min Parent-and-Children [87], the Hill Climbing method [87] or the Max-Min Hill-Climbing heuristic [87], produced structures that were very similar to the Gaussian Naive Bayes configuration, but in reverse (all arrows went from the outcome to other features). Such structures did not fit outcome prediction and they did not allow to retrieve any additional information from feature sets. The Structural Expectation-Maximization algorithm [88], [89] also failed to achieve meaningful structures. Although they were different from the structures obtained with three previous methods, there were some clear problems. For example, Figure 6.1 shows one such structure obtained using the *T1+T2 (5, RF)* feature set. According



(A) (B)

FIGURE 6.1: The Bayesian Network structure of the *T1+T2 (5, RF)* feature set achieved with the Structural Expectation-Maximization algorithm

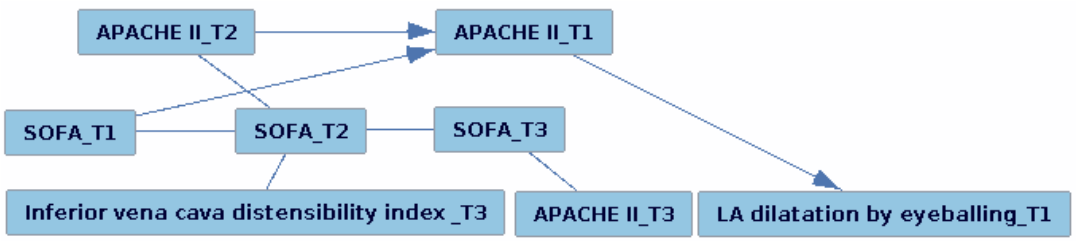
to this structure, the *Lactate levels* at time-step 2 influence the *Respiratory rate* at time-step 1, which is clearly wrong. And since all the directions are clear, it is very hard to explain such connectivity between the features. Additionally, from the same figure it can be seen that after adding the target feature, the CBN substantially changes. This happens with all other CBNs that are built with this method, meaning that that the CBNs are unstable.

The problems with the aforementioned algorithms can be caused by the small size of the dataset. In order to address this issue, it was decided to use the Fast Greedy Search (FGES) algorithm. It also returns the most probable CNB, but should it not be able to determine the direct causation, it would return an undirected edge between two features. So, technically, the algorithm returns a "pattern" containing arcs, which represent direct causation, and undirected edges, where such an edge indicates there is a causal arc, but its direction cannot be determined. Additionally, it works with real-valued variables, and most of the features in this chapter are of that type.

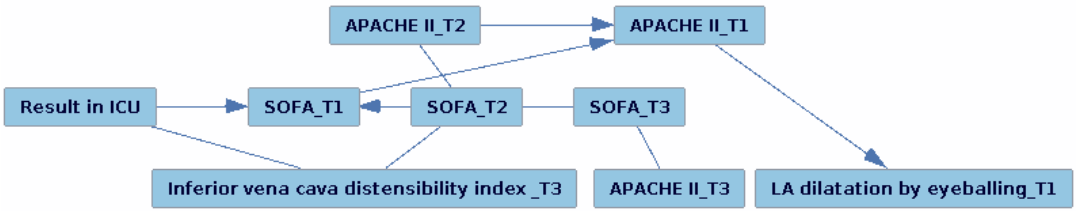
The algorithm was applied twice for each of the feature sets obtained in the previous chapter. First, the algorithm was applied only to the features, excluding the outcome, and the second time it was applied to both features and the outcome. The expectation was to have few or no differences between features that did not involve the outcome. That would mean that the CBNs are stable enough to draw some clear conclusions.

6.1.1 FGES Algorithm

The Fast Greedy Search (FGES) algorithm is an optimized and parallelized version of the Greedy Equivalence Search algorithm [90]. It heuristically searches the space of CBNs and returns the model with the highest Bayesian Information Criterion (BIC)



(A)



(B)

FIGURE 6.2: The CBN for the *Full+ (9, UFS)* feature set achieved with: a) without the target feature; 2) with the target feature.

[91] it finds. It starts with the empty graph and then adds new edges to increase the Bayesian score. The process stops when there is no edges that increase the score. After that, it starts removing the edges until no single edge removal can increase the score.

The algorithm works on the assumption that the causal process generating the data is accurately modeled by a CBN. Each node in this CBN is a a linear function of its parents, plus a finite additive Gaussian noise term. Each observation in the data is independent, and was obtained by randomly sampling all the variables from the joint distribution. Given all these assumptions, the FGES procedure outputs the CBN structure that contains:

1. an arc X $\rightarrow$ Y, if and only if X causes Y;
2. an edge (—), if and only if either X causes Y or Y causes X;
3. no edge between X and Y, if and only if X and Y have no direct causal relationship between them.

The FGES algorithm was used with the help of the *Tetrad* software (see `http://www.phil.cmu.edu/tetrad/`).

6.1.2 Results

The full lists of features for feature sets in this section can be found at the end of the previous chapter (see Listing 5.3). Here, for each of the feature sets there are two corresponding CBNs. The first one (the CBN "a") was built without the target feature (*Result in ICU*), the second one (the CBN "b") included all features and the target.

Figure 6.2 shows the CBNs for the *Full+ (9, UFS)* feature set. After adding the target feature, the direction of the edge between *SOFA* at *T1* and *T2* changed. According to the second CBN, the *Result in ICU* depends only on *Inferior vena cava distensibility index* at *T3*. Both CBNs are fully connected and have complex interconnectivity.

Figure 6.3 shows the CBNs for the *Full+ (9, UFS)* feature set. The difference between the previous feature set and this one is the addition of *X Norepinephrine* at

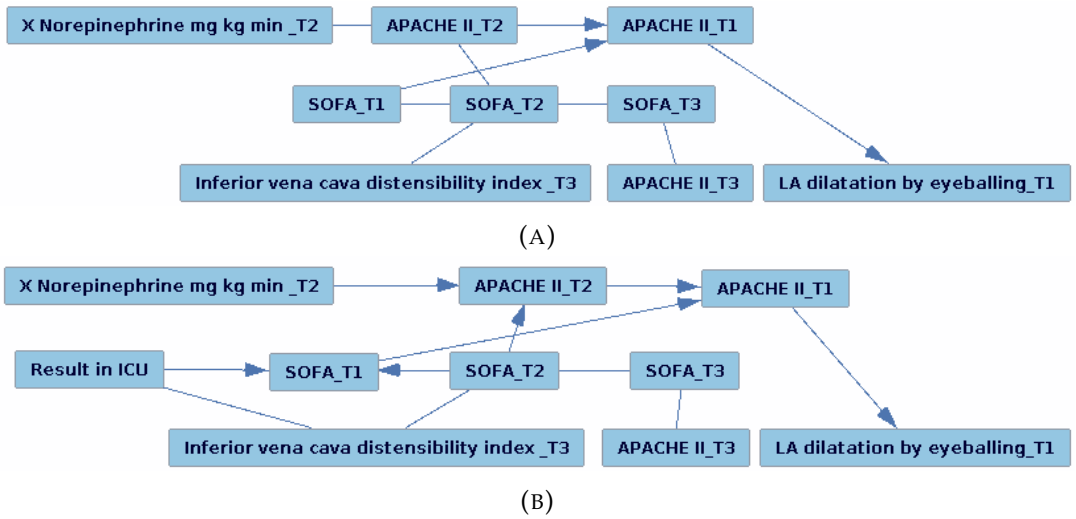


FIGURE 6.3: The CBN for the *Full+ (10, UFS)* feature set achieved with: a) without the target feature; 2) with the target feature.

T2. According to the CBNs, it only influences *APACHE II* at *T2* and after adding the target feature the direction of the causality becomes clear. There are no other differences for this feature set, as compared to the previous one, meaning that the CBNs for both of the feature sets are stable.

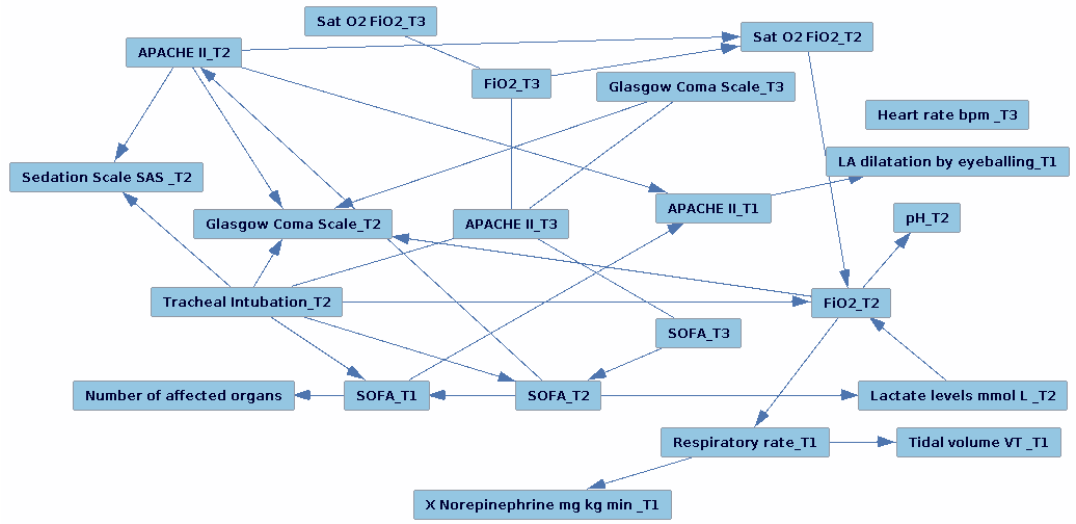
Figure 6.4 shows the CBNs for the *Full (24, UFS)* feature set. The edges in two CBNs are exactly the same (except the ones that are connected to the target value). The target feature in the second CBN is connected to *FiO2* at *T2*, *SOFA* at *T2* and *X Norepinephrine* at *T1*. This is the biggest feature set in the experiments and it is very interesting to compare the connections here to those of the rest of the feature sets. A lot of the directions in the edges are missing, the node *Heart rate* at *T3* is disconnected from the network. Finally, *SOFA* at *T2* is the only direct influence on the target.

Figure 6.5 shows the CBNs for the *Full (7, RF)* feature set. Both CBNs do not have any edges with the established directions of causality. Once again, *Heart rate* at *T3* is disconnected from the rest of the network. This time there is a second node that do not have any edges which is *Respiratory rate* at *T1*. After adding the outcome, only one new undirected edge between *SOFA* at *T2* and *Result in ICU* appeared. This is the second feature set, where *SOFA* at *T2* is connected to the outcome.

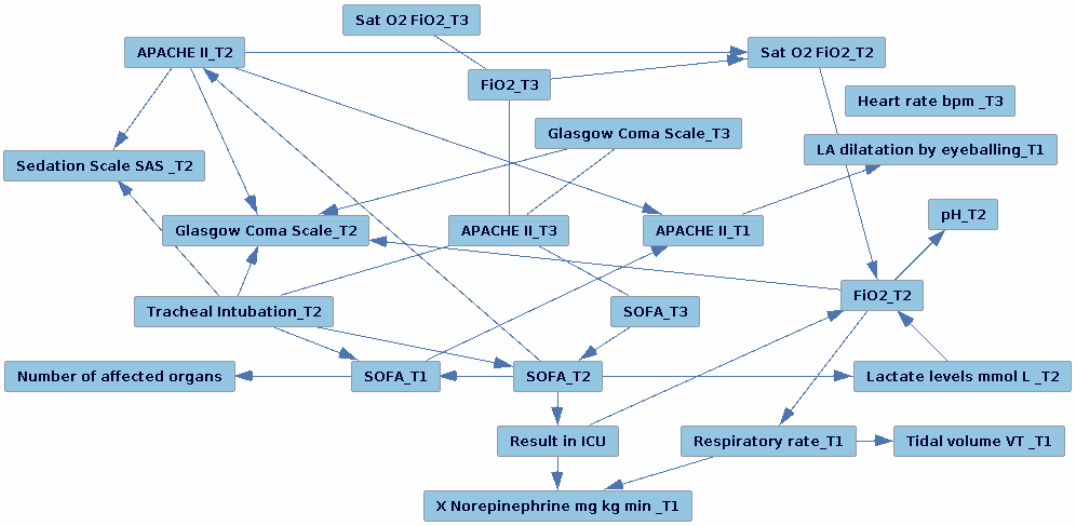
Figure 6.5 shows the CBNs for the *Full (15, RF)* feature set. In these CBNs there are two disconnected nodes for *Heart rate* at *T3* and *E wave* at *T1* features. After adding the target feature, the CBN changes slightly. The directed edge from *APACHE II* at *T1* to *X Norepinephrine* at *T1* disappeared, and a new undirected connection between *X Norepinephrine* at *T1* and *SOFA* at *T2* was added. The edge from *X norepinephrine* at *T1* to *Respiratory rate* at *T1* became undirected. Once again, *SOFA* at *T2* is the only direct connection to the outcome.

Figure 6.7 shows the CBNs for the *T1+T2 (5, RF)* feature set. This is the smallest feature set in the experiments. And two nodes in both CBNs are disconnected: *E wave* at *T1* and *Respiratory rate* at *T1*. Here, *SOFA* at *T2* is the only direct connection for the target feature. The connections between the feature are very similar to those observed in the previous feature sets.

In all pairs of CBNs, the presence of the target value did not generate much change. In some cases, new edges appeared, but, in the majority of cases, only few edges changed their directions. It should be noted that the majority of edges were

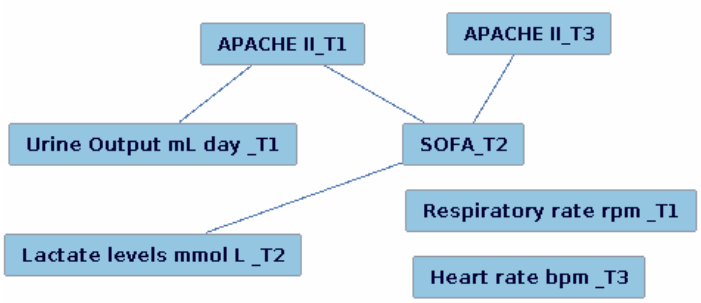


(A)

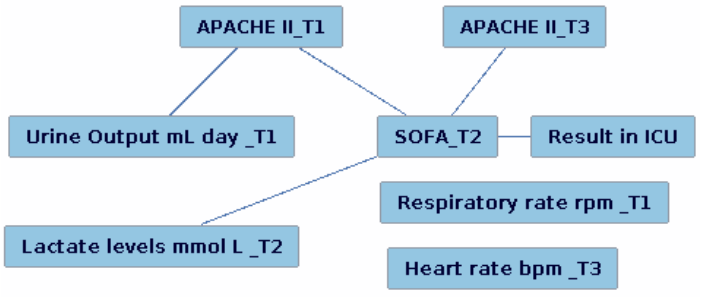


(B)

FIGURE 6.4: The CBN for the *Full (24, UFS)* feature set achieved with: a) without the target feature; 2) with the target feature.



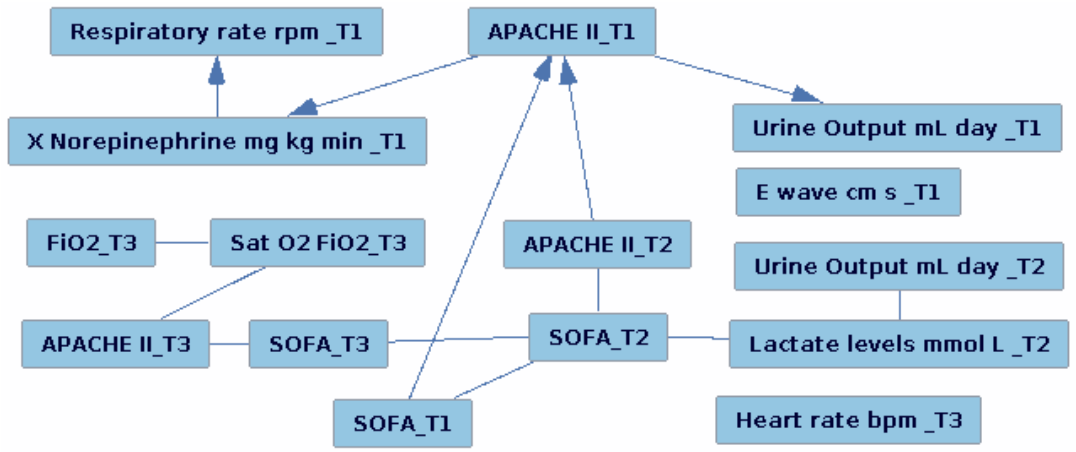
(A)



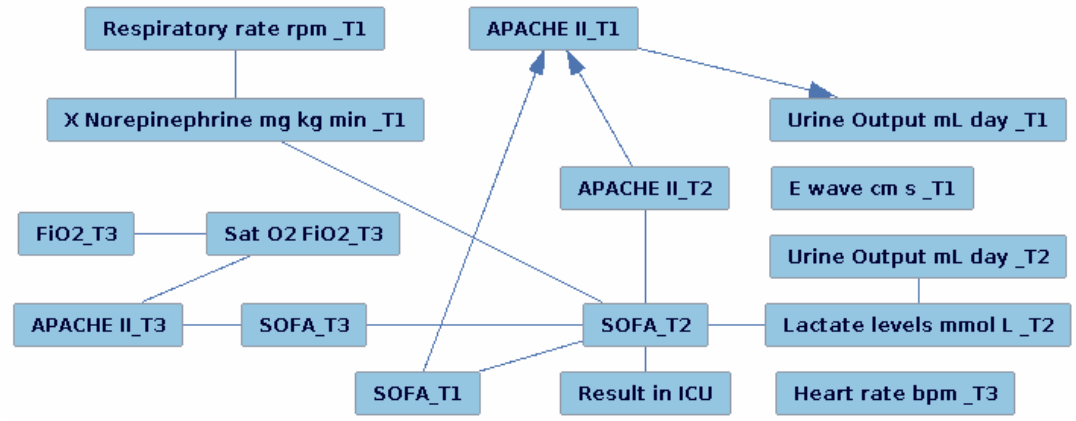
(B)

FIGURE 6.5: The CBN for the *Full (7, RF)* feature set achieved with: a) without the target feature; 2) with the target feature.

undirected. This might be because of the small amount of data available for model generation. Nevertheless, some clear patterns are noticeable throughout the experiments. The next chapter summarizes the results and draws the conclusions of the thesis work.

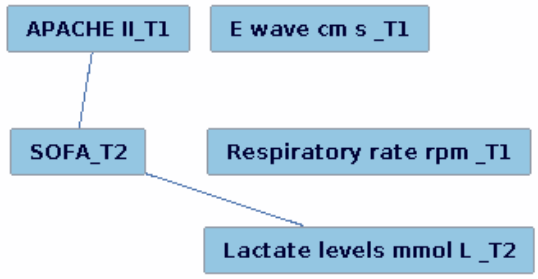


(A)

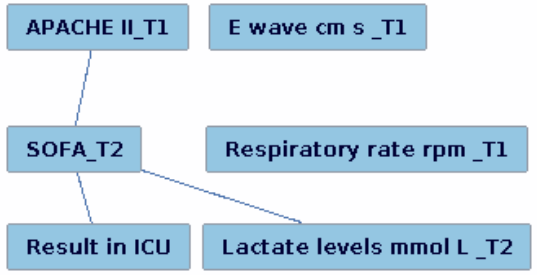


(B)

FIGURE 6.6: The CBN for the *Full (15, RF)* feature set achieved with: a) without the target feature; 2) with the target feature.



(A)



(B)

FIGURE 6.7: The CBN for the *T1+T2 (5, RF)* feature set achieved with: a) without the target feature; 2) with the target feature.

Chapter 7

Discussion and Conclusions

The previous chapters introduced the problem of mortality prediction in patients with shock. They described shock pathology, the biomarkers that are currently used to detect it and different approaches that allow to predict outcome of diseases. They also introduced three major experiment sets with their methodologies and results. Here, in this chapter the results of the experiments are discussed to draw a number of conclusions. Apart from conclusions this chapter addresses the limitations of the thesis and proposes possible future lines of work.

7.1 Conclusions

The work described in this thesis has led to a list of the biomarkers that can be potentially helpful for the outcome prediction in patients with shock. The work identified that, at different time steps, some features might be more or less important than others in those terms. The result is a list of promising features that can be found in Listing 7.1, with the corresponding time steps when these features are relevant ($T1$ < 16 hours from the diagnosis of shock, $T2$ = 48 h after the diagnosis; $T3$ = day 7 since the diagnosis). The subsection discusses conclusions of the experiments in more details.

```
Features from the Causal Discovery experiments: SOFA T1, SOFA T2, SOFA
    T3, APACHE II T1, APACHE II T2, Inferior vena cava distensibility
    index T3, Lactate Levels T2, Tracheal Intubation T2, X
    Norepinephrine T1, FiO2 at T2;

Additional features from the Feature Selection experiments: Respiratory
     rate T1, Urine Output T1, Urine Output T2, Heart rate T3.
```

List of biomarkers that were considered promising for the outcome prediction in patients with shock (see Sections 7.1.2 and 7.1.3 for more details).

7.1.1 Preliminary Experiments

The first in the series of experiments were the preliminary experiments from Chapter 4. Their main goal was to establish the missing data imputation technique and the ML model that performed best for mortality prediction task. The Gaussian Naive Bayes turned out to be such best model. As for the imputation technique, imputation by linear regression through prediction showed the best results (the lowest reconstruction error). However, since it only worked with the numerical features, the Random Forest imputation was used for all later experiments instead. It was the imputation method with the lowest reconstruction error.

In addition to justifying the choice of the imputation technique for the later usage, the preliminary experiments showed the feasibility of using Bayesian Networks for mortality prediction. Bayesian Networks showed a comparable performance to other tested ML models in the analyzed datasets. This is an important result, since these models provide a certain level of interpretability to the outcome prediction through probabilities and the structure of the network. And interpretable models allow knowledge extraction and improve understanding of causal relationships between the features. Finally, the tested structures for the Bayesian Networks raise a doubt that *SOFA* and *APACHE II* at the time-step 3 are helpful for the mortality prediction, at least in the analyzed datasets. The models generated without them showed much better performance.

7.1.2 Feature Selection Experiments

The second group of experiments was dedicated to the feature selection procedure in four subsets of data: **Full+**, **Full**, **T1+T2**, **T1**. It was described in Chapter 5. The goal of these experiments was to identify promising feature sets. The criteria for their selection was the performance of the Gaussian Naive Bayes model with these feature sets. After all feature sets were filtered based on the performance, only six of them left (see Listing 5.3). Apart from this, stability scores were calculated and they were used as measures of usefulness for individual features in the resulted feature sets.

By looking at stability scores, one may notice that the RF stability scores have a lot of reoccurring features, while the scores for other feature selection methods are very diverse. The later causal discover experiments show that the RF scores were more accurate. The features like *SOFA*, *APACHE II*, *Lactate levels*, *Heart rate*, *Respiratory rate*, *Urine Output* and *X Norepinephrine* usually scored high. In most of the cases, *SOFA* and *APACHE II* features were also highly rated by other feature selection methods.

Although such patterns of features are noticeable, the differences between feature sets in different datasets are also important. For example, if we compare the stability scores of the **Full+** and the **Full** datasets, we observe that after removing the features with more than 50% of missing values, there are a lot of similarities between the first few best features for the RF and other selection methods. Both stability scores rank *SOFA*, *APACHE II* and *Heart rate* high. Interestingly, there is a gain to the performance after removing the *T3* features, whereas removing both the *T2* and *T3* features considerably worsens the performance. It may occur because *T3* features have a lot of missing values, and basing the outcome prediction on *T1* features is not sufficient.

Additionally, it seems that certain features are highly valuable only at certain time steps. For example, *Lactate levels* seems to have the most predictive value at *T2*, *X Norepinephrine* and *Respiratory rate* are particularly valuable at *T1*, *Urine Output* – at *T1* and *T2* and *Heart rate* at *T3*. These conclusions are based on the performances of the individual feature sets for the outcome prediction and their stability scores. These hypotheses were further tested in the causal discovery experiments.

7.1.3 Causal Discovery Experiments

The causal discovery experiments are described in Chapter 6. The purpose of these experiments was to identify the causal relationships of biomarkers and the outcome. The Bayesian Network approach was used to construct two structures for each of the six feature sets. The first one was built only with features without the outcome

feature, and the second one used all of them. The experiments showed that the structures were stable enough and adding the target feature did not change much in the causal relationships. The causal Bayesian Networks structures revealed the connections between the features and the outcome.

The assumption behind the evaluation of the most important biomarkers for the outcome prediction was that such features are the closest to the outcome in a graph. The direction of the edge from feature to the outcome is also a good indicator that the feature is important. According to six structures that contain *Result in ICU*, *SOFA* at the time-step 2 is the most important feature for the outcome prediction. In almost all structure it has the direct connection to the outcome. The features that are connected to *SOFA* at *T2* are also valuable. The CBNs for the *Full+ (9, UFS)* and the *Full+ (10, UFS)* feature sets indicate the importance of *APACHE II* at *T2*, *SOFA* at *T3*, *Inferior vena cava distensibility index* at *T3*. The CBNs for the *Full (24, UFS)* feature set add *SOFA* at *T1*, *Lactate levels* at *T2*, *Tracheal Intubation* at *T2* to this list. In these CBNs the outcome influences *X Norepinephrine* at *T1* and *FiO2* at *T2*, and they might be also important indicators for the outcome prediction. The rest of the CBNs for the last three feature sets confirm some of the candidates for mortality indicators. They confirm the usage of *SOFA* at *T1*, *T2* and *T3*, *APACHE II* at T2, *Lactate Levels* at *T2* and *X Norepinephrine* at *T1*. The CBNs for the *Full (7, RF)* and the *T1+T2 (5, RF)* feature sets also justify the usefulness of *APACHE II* at *T1*.

These findings agree with other studies that support the use of certain biomarkers for mortality prediction. For instance, the SOFA score shows a significant prognostic value for in-hospital mortality prediction [37] [36]. The APACHE II score is often used for the outcome prediction too [39]–[41]. And in the study by Jansen *et al.* [22], targeting a decrease of at least 20% in the blood lactate level over a 2-hour period seemed to be associated with reduced in-hospital mortality in patients with shock. However, some of the biomarkers that are also related to the mortality in patients with shock, like mean blood pressure [13], mean arterial pressure [92], or low LVEF [14] did not show significant predictive value in this study.

7.2 Criticism and Future Work

The study at hand has several limitations. Firstly, the analyzed dataset has too few observations and too many missing data. This means that the findings of this study might not be general enough. And in order to increase the confidence in the assumption that the better the performance of a certain feature set, the more important individual features, more cases should be added to the study. Some of the patterns might have been missed because there were too few cases that represent them, and some patterns that were considered general enough might turn out to be very specific for the patients in the dataset. Additionally, some of the features, that may be good indicators for mortality, might have a lot of missing values and lost valuable information for prediction. In order to address these issues, the findings of this study should be tested on a different and considerably larger dataset.

Secondly, the study attempted to identify mortality prediction indicators for circulatory shock. Unfortunately, the dataset contained only septic and cardiogenic shock patients with a control group. So naturally, the biomarkers that were found in this study should be tested for hypovolaemic and obstructive shock, as well. Septic and cardiogenic shocks are the most common types, but the treatment and the diagnosis of the shock type ideally should go in parallel. And if two other types of shock can have completely different biomarkers for mortality prediction, then this

makes it difficult to plan the treatment. This problem can be addressed by including patients with other two types of shock and repeating the feature selection and causal discovery experiments.

Third, although the study proposed the causal relationships for the promising feature sets, there is still a lot of missing relationship information. Most of the edges in the Bayesian Network structures are undirected and hence, it is impossible to build a model that can be tested and evaluated. The uncertainties in the structures are very likely related to a small number of observation. So in order to build the Bayesian Network that could predict the outcome of a patient with shock, these uncertainties should be eliminated. Once the structures are known, it is required to measure the performance of the model. And the training and the testing should be done using the dataset with enough observations

The work of this Master's thesis was based on the ShockOmics project dataset. Therefore, the criticism and the future work proposed in this section are closely related to improving the quality of data. Although the ShockOmics dataset proposes a huge number of clinical features, more patient observations and more types of shock are needed to corroborate the findings of the thesis and build an interpretable model.

Bibliography

[1] J.-L. Vincent and D. De Backer, "Circulatory shock", *New England Journal of Medicine*, vol. 369, no. 18, pp. 1726–1734, 2013.

[2] S. A. Sterling, M. A. Puskarich, A. F. Glass, F. Guirgis, and A. E. Jones, "The impact of the sepsis-3 septic shock definition on previously defined septic shock patients", *Critical care medicine*, vol. 45, no. 9, pp. 1436–1442, 2017.

[3] H. L. Nguyen, J. Yarzebski, D. Lessard, J. M. Gore, D. D. McManus, and R. J. Goldberg, "Ten-year (2001–2011) trends in the incidence rates and short-term outcomes of early versus late onset cardiogenic shock after hospitalization for acute myocardial infarction", *Journal of the American Heart Association*, vol. 6, no. 6, e005566, 2017.

[4] R. McNally, P Mair, B. Mugno, and B. Riemann, "Co-morbid obsessive–compulsive disorder and depression: A bayesian network approach", *Psychological medicine*, vol. 47, no. 7, pp. 1204–1214, 2017.

[5] Y. Luo, I. El Naqa, D. L. McShan, D. Ray, I. Lohse, M. M. Matuszak, D. Owen, S. Jolly, T. S. Lawrence, R. K. Ten Haken, *et al.*, "Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via bayesian network analysis", *Radiotherapy and Oncology*, vol. 123, no. 1, pp. 85–92, 2017.

[6] G. Cooper, C. Cai, and X. Lu, "Tumor-specific causal inference (tci): A bayesian method for identifying causative genome alterations within individual tumors", *BioRxiv*, p. 225 631, 2018.

[7] M. Cecconi, D. De Backer, M. Antonelli, R. Beale, J. Bakker, C. Hofer, R. Jaeschke, A. Mebazaa, M. R. Pinsky, J. L. Teboul, *et al.*, "Consensus on circulatory shock and hemodynamic monitoring. task force of the european society of intensive care medicine", *Intensive care medicine*, vol. 40, no. 12, pp. 1795–1815, 2014.

[8] T. S. O. Nunes, R. T. Ladeira, A. T. Bafi, L. C. P. de Azevedo, F. R. Machado, and F. G. R. Freitas, "Duration of hemodynamic effects of crystalloids in patients with circulatory shock after initial resuscitation", *Annals of intensive care*, vol. 4, no. 1, p. 25, 2014.

[9] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, *et al.*, "The third international consensus definitions for sepsis and septic shock (sepsis-3)", *Jama*, vol. 315, no. 8, pp. 801–810, 2016.

[10] F. Aletti, C. Conti, M. Ferrario, V. Ribas, B. B. Pinto, A. Herpain, E. Post, E. R. Medina, C. Barlassina, E. de Oliveira, *et al.*, "Shockomics: Multiscale approach to the identification of molecular biomarkers in acute heart failure induced by shock", *Scandinavian journal of trauma, resuscitation and emergency medicine*, vol. 24, no. 1, p. 9, 2016.

[11] A. M. Schulman, J. A. Claridge, G. Carr, D. L. Diesen, and J. S. Young, "Predictors of patients who will develop prolonged occult hypoperfusion following blunt trauma", *Journal of Trauma and Acute Care Surgery*, vol. 57, no. 4, pp. 795–800, 2004.

[12] A. Babaev, P. D. Frederick, D. J. Pasta, N. Every, T. Sichrovsky, J. S. Hochman, N. Investigators, *et al.*, "Trends in management and outcomes of patients with acute myocardial infarction complicated by cardiogenic shock", *Jama*, vol. 294, no. 4, pp. 448–454, 2005.

[13] M. Shankar-Hari, G. S. Phillips, M. L. Levy, C. W. Seymour, V. X. Liu, C. S. Deutschman, D. C. Angus, G. D. Rubenfeld, and M. Singer, "Developing a new definition and assessing new clinical criteria for septic shock: For the third international consensus definitions for sepsis and septic shock (sepsis-3)", *Jama*, vol. 315, no. 8, pp. 775–787, 2016.

[14] V.-P. Harjola, J. Lassus, A. Sionis, L. Køber, T. Tarvasmäki, J. Spinar, J. Parissis, M. Banaszewski, J. Silva-Cardoso, V. Carubelli, *et al.*, "Clinical picture and risk prediction of short-term mortality in cardiogenic shock", *European journal of heart failure*, vol. 17, no. 5, pp. 501–509, 2015.

[15] Y. Sakr, K. Reinhart, J.-L. Vincent, C. L. Sprung, R. Moreno, V. M. Ranieri, D. De Backer, and D. Payen, "Does dopamine administration in shock influence outcome? results of the sepsis occurrence in acutely ill patients (soap) study", *Critical care medicine*, vol. 34, no. 3, pp. 589–597, 2006.

[16] D. De Backer, P. Biston, J. Devriendt, C. Madl, D. Chochrad, C. Aldecoa, A. Brasseur, P. Defrance, P. Gottignies, and J.-L. Vincent, "Comparison of dopamine and norepinephrine in the treatment of shock", *New England Journal of Medicine*, vol. 362, no. 9, pp. 779–789, 2010.

[17] C. Fleischmann, D. O. Thomas-Rueddel, M. Hartmann, C. S. Hartog, T. Welte, S. Heublein, U. Dennler, and K. Reinhart, "Hospital incidence and mortality rates of sepsis: An analysis of hospital episode (drg) statistics in germany from 2007 to 2013", *Deutsches Ärzteblatt International*, vol. 113, no. 10, p. 159, 2016.

[18] J.-P. Quenot, C. Binquet, F. Kara, O. Martinet, F. Ganster, J.-C. Navellou, V. Castelain, D. Barraud, J. Cousson, G. Louis, *et al.*, "The epidemiology of septic shock in french intensive care units: The prospective multicenter cohort episs study", *Critical care*, vol. 17, no. 2, R65, 2013.

[19] M. Antonelli, M. Levy, P. J. Andrews, J. Chastre, L. D. Hudson, C. Manthous, G. U. Meduri, R. P. Moreno, C. Putensen, T. Stewart, *et al.*, "Hemodynamic monitoring in shock and implications for management", *Intensive care medicine*, vol. 33, no. 4, pp. 575–590, 2007.

[20] L. De Luca, Z. Olivari, A. Farina, L. Gonzini, D. Lucci, A. Di Chiara, G. Casella, F. Chiarella, A. Boccanelli, G. Di Pasquale, *et al.*, "Temporal trends in the epidemiology, management, and outcome of patients with cardiogenic shock complicating acute coronary syndromes", *European journal of heart failure*, vol. 17, no. 11, pp. 1124–1132, 2015.

[21] R. M. Otero, H. B. Nguyen, D. T. Huang, D. F. Gaieski, M. Goyal, K. J. Gunnerson, S. Trzeciak, R. Sherwin, C. V. Holthaus, T. Osborn, *et al.*, "Early goal-directed therapy in severe sepsis and septic shock revisited: Concepts, controversies, and contemporary findings", *Chest*, vol. 130, no. 5, pp. 1579–1595, 2006.

[22] T. C. Jansen, J. van Bommel, F. J. Schoonderbeek, S. J. Sleeswijk Visser, J. M. van der Klooster, A. P. Lima, S. P. Willemsen, and J. Bakker, "Early lactate-guided therapy in intensive care unit patients: A multicenter, open-label, randomized controlled trial", *American journal of respiratory and critical care medicine*, vol. 182, no. 6, pp. 752–761, 2010.

[23] D. Kim, J.-G. Joung, K.-A. Sohn, H. Shin, Y. R. Park, M. D. Ritchie, and J. H. Kim, "Knowledge boosting: A graph-based integration approach with multi-omics data and genomic knowledge for cancer clinical outcome prediction", *Journal of the American Medical Informatics Association*, vol. 22, no. 1, pp. 109–120, 2014.

[24] K. Cisek, M. Krochmal, J. Klein, and H. Mischak, "The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease", *Nephrology Dialysis Transplantation*, vol. 31, no. 12, pp. 2003–2011, 2015.

[25] S. Klau, V. Jurinovic, R. Hornung, T. Herold, and A.-L. Boulesteix, "Priority-lasso: A simple hierarchical approach to the prediction of clinical outcome using multi-omics data", 2017.

[26] G. M. Douglas, R. Hansen, C. M. Jones, K. A. Dunn, A. M. Comeau, J. P. Bielawski, R. Tayler, E. M. El-Omar, R. K. Russell, G. L. Hold, *et al.*, "Multi-omics differentially classify disease state and treatment outcome in pediatric crohn's disease", *Microbiome*, vol. 6, no. 1, p. 13, 2018.

[27] D. Kim, R. Li, A. Lucas, S. S. Verma, S. M. Dudek, and M. D. Ritchie, "Using knowledge-driven genomic interactions for multi-omics data analysis: Metadimensional models for predicting clinical outcomes in ovarian carcinoma", *Journal of the American Medical Informatics Association*, vol. 24, no. 3, pp. 577–587, 2016.

[28] M. Ferrario, A. Cambiaghi, L. Brunelli, S. Giordano, P. Caironi, L. Guatteri, F. Raimondi, L. Gattinoni, R. Latini, S. Masson, *et al.*, "Mortality prediction in patients with severe septic shock: A pilot study using a target metabolomics approach", *Scientific reports*, vol. 6, p. 20 391, 2016.

[29] V. Pettilä, M. Pettilä, S. Sarna, P. Voutilainen, and O. Takkunen, "Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill", *Critical Care Medicine*, vol. 30, no. 8, pp. 1705–1711, 2002.

[30] J.-R. Le Gall, "The use of severity scores in the intensive care unit", *Intensive care medicine*, vol. 31, no. 12, pp. 1618–1623, 2005.

[31] A. G. Rapsang and D. C. Shyam, "Scoring systems in the intensive care unit: A compendium", *Indian journal of critical care medicine: Peer-reviewed, official publication of Indian Society of Critical Care Medicine*, vol. 18, no. 4, p. 220, 2014.

[32] H. Champion, "Trauma scoring", *Scandinavian Journal of Surgery*, vol. 91, no. 1, pp. 12–22, 2002.

[33] J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. Reinhart, P. Suter, and L. Thijs, *The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure*, 1996.

[34] J.-L. Vincent, A. De Mendonça, F. Cantraine, R. Moreno, J. Takala, P. M. Suter, C. L. Sprung, F. Colardyn, and S. Blecher, "Use of the sofa score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicenter, prospective study", *Critical care medicine*, vol. 26, no. 11, pp. 1793–1800, 1998.

[35] F. L. Ferreira, D. P. Bota, A. Bross, C. Mélot, and J.-L. Vincent, "Serial evaluation of the sofa score to predict outcome in critically ill patients", *Jama*, vol. 286, no. 14, pp. 1754–1758, 2001.

[36] E. P. Raith, A. A. Udy, M. Bailey, S. McGloughlin, C. MacIsaac, R. Bellomo, and D. V. Pilcher, "Prognostic accuracy of the sofa score, sirs criteria, and qsofa score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit", *Jama*, vol. 317, no. 3, pp. 290–300, 2017.

[37] R. Rivera-Fernández, R. Nap, G. Vázquez-Mata, and D. R. Miranda, "Analysis of physiologic alterations in intensive care unit patients and their relationship with mortality", *Journal of critical care*, vol. 22, no. 2, pp. 120–128, 2007.

[38] W. A. Knaus, E. A. Draper, D. P. Wagner, and J. E. Zimmerman, "Apache ii: A severity of disease classification system.", *Critical care medicine*, vol. 13, no. 10, pp. 818–829, 1985.

[39] Y. Huang, J. Chen, S. Zhong, and J. Yuan, "Role of apache ii scoring system in the prediction of severity and outcome of acute intracerebral hemorrhage", *International Journal of Neuroscience*, vol. 126, no. 11, pp. 1020–1024, 2016.

[40] A. Agarwal, G. S. Choudhary, M. Bairwa, and A. Choudhary, "Apache ii scoring in predicting surgical outcome in patients of perforation peritonitis", *International Surgery Journal*, vol. 4, no. 7, pp. 2321–2325, 2017.

[41] H Lee, C. Lim, H. Hong, J. Ju, Y. Jeon, J. Hwang, and H. Park, "Efficacy of the apache ii score at icu discharge in predicting post-icu mortality and icu readmission in critically ill surgical patients.", *Anaesthesia & Intensive Care*, vol. 43, no. 2, 2015.

[42] R. Raj, J. Siironen, R. Kivisaari, J. Hernesniemi, and M. B. Skrifvars, "Predicting outcome after traumatic brain injury: Development of prognostic scores based on the impact and the apache ii", *Journal of neurotrauma*, vol. 31, no. 20, pp. 1721–1732, 2014.

[43] G. Teasdale and B. Jennett, "Assessment of coma and impaired consciousness: A practical scale", *The Lancet*, vol. 304, no. 7872, pp. 81–84, 1974.

[44] E Cantais, O Paut, R Giorgi, L Viard, and J Camboulives, "Evaluating the prognosis of multiple, severely traumatized children in the intensive care unit", *Intensive care medicine*, vol. 27, no. 9, pp. 1511–1517, 2001.

[45] P. Emami, P. Czorlich, F. S. Fritzsche, M. Westphal, J. M. Rueger, R. Lefering, and M. Hoffmann, "Impact of glasgow coma scale score and pupil parameters on mortality rate and outcome in pediatric and adult severe traumatic brain injury: A retrospective, multicenter cohort study", *Journal of neurosurgery*, vol. 126, no. 3, pp. 760–767, 2017.

[46] D. H. Fulkerson, I. K. White, J. M. Rees, M. M. Baumanis, J. L. Smith, L. L. Ackerman, J. C. Boaz, and T. G. Luerssen, "Analysis of long-term (median 10.5 years) outcomes in children presenting with traumatic brain injury and an initial glasgow coma scale score of 3 or 4", *Journal of Neurosurgery: Pediatrics*, vol. 16, no. 4, pp. 410–419, 2015.

[47] P. G. Bastos, X. Sun, D. P. Wagner, A. W. Wu, and W. A. Knaus, "Glasgow coma scale score in the evaluation of outcome in the intensive care unit: Findings from the acute physiology and chronic health evaluation iii study.", *Critical care medicine*, vol. 21, no. 10, pp. 1459–1465, 1993.

[48] E. F. Wijdicks, A. A. Kramer, T. Rohs Jr, S. Hanna, F. Sadaka, J. O'Brien, S. Bible, S. M. Dickess, and M. Foss, "Comparison of the full outline of unresponsiveness score and the glasgow coma scale in predicting mortality in critically ill patients", *Critical care medicine*, vol. 43, no. 2, pp. 439–444, 2015.

[49] E. F. Wijdicks, A. A. Rabinstein, W. R. Bamlet, and J. N. Mandrekar, "Four score and glasgow coma scale in predicting outcome of comatose patients: A pooled analysis", *Neurology*, vol. 77, no. 1, pp. 84–85, 2011.

[50] S Chandrasekhar, M. A. Rahim, S. Quraishi, C. R. Theja, and K. S. Kiran, "An observational clinical study of assessing the utility of pss (poison severity score) and gcs (glasgow coma scale) scoring systems in predicting severity and clinical outcomes in op poisoning", *Journal of Evidence Based Medicine and Healthcare*, vol. 4, no. 38, pp. 2325–2332, 2017.

[51] N. Naeem, "Beyond the intensive care unit: A review of interventions aimed at anticipating and preventing in-hospital cardiopulmonary arrest", *Resuscitation*, vol. 73, no. 3, pp. 491–492, 2007.

[52] M. R. Pinsky and A. Dubrawski, "Gleaning knowledge from data in the intensive care unit", *American journal of respiratory and critical care medicine*, vol. 190, no. 6, pp. 606–610, 2014.

[53] M. Motwani, D. Dey, D. S. Berman, G. Germano, S. Achenbach, M. H. Al-Mallah, D. Andreini, M. J. Budoff, F. Cademartiri, T. Q. Callister, *et al.*, "Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis", *European heart journal*, vol. 38, no. 7, pp. 500–507, 2016.

[54] J. Allyn, N. Allou, P. Augustin, I. Philip, O. Martinet, M. Belghiti, S. Provenchere, P. Montravers, and C. Ferdynus, "A comparison of a machine learning model with euroscore ii in predicting mortality after elective cardiac surgery: A decision curve analysis", *PloS one*, vol. 12, no. 1, e0169772, 2017.

[55] M. M. Churpek, T. C. Yuen, C. Winslow, D. O. Meltzer, M. W. Kattan, and D. P. Edelson, "Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards", *Critical care medicine*, vol. 44, no. 2, p. 368, 2016.

[56] R. A. Taylor, J. R. Pare, A. K. Venkatesh, H. Mowafi, E. R. Melnick, W. Fleischman, and M. K. Hall, "Prediction of in-hospital mortality in emergency department patients with sepsis: A local big data–driven, machine learning approach", *Academic emergency medicine*, vol. 23, no. 3, pp. 269–278, 2016.

[57] N. T. Liu, J. B. Holcomb, C. E. Wade, M. I. Darrah, and J. Salinas, "Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for lifesaving interventions in trauma patients", *Shock*, vol. 42, no. 2, pp. 108–114, 2014.

[58] J. D. Frizzell, L. Liang, P. J. Schulte, C. W. Yancy, P. A. Heidenreich, A. F. Hernandez, D. L. Bhatt, G. C. Fonarow, and W. K. Laskey, "Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: Comparison of machine learning and other statistical approaches", *Jama cardiology*, vol. 2, no. 2, pp. 204–209, 2017.

[59] A.-L. Boulesteix, R. De Bin, X. Jiang, and M. Fuchs, "Ipf-lasso: Integrative-penalized regression with penalty factors for prediction based on multi-omics data", *Computational and mathematical methods in medicine*, vol. 2017, 2017.

[60] Q. Zhao, X. Shi, Y. Xie, J. Huang, B. Shia, and S. Ma, "Combining multidimensional genomic measurements for predicting cancer prognosis: Observations from tcga", *Briefings in bioinformatics*, vol. 16, no. 2, pp. 291–303, 2014.

[61] M. Fuchs, T. Beißbarth, E. Wingender, and K. Jung, "Connecting high-dimensional mrna and mirna expression data for binary medical classification problems", *Computer methods and programs in biomedicine*, vol. 111, no. 3, pp. 592–601, 2013.

[62] A. Liaw, M. Wiener, *et al.*, "Classification and regression by randomforest", *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[63] R. Tibshirani, "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[64] M. Kim, N. Rai, V. Zorraquino, and I. Tagkopoulos, "Multi-omics integration accurately predicts cellular state in unexplored conditions for escherichia coli", *Nature communications*, vol. 7, p. 13 090, 2016.

[65] R. Shouval, M. Labopin, O. Bondi, H. Mishan-Shamay, A. Shimoni, F. Ciceri, J. Esteve, S. Giebel, N. C. Gorin, C. Schmid, *et al.*, "Prediction of allogeneic hematopoietic stem-cell transplantation mortality 100 days after transplantation using a machine learning algorithm: A european group for blood and marrow transplantation acute leukemia working party retrospective data mining study", *Journal of Clinical Oncology*, vol. 33, no. 28, pp. 3144–3151, 2015.

[66] H.-Y. Lu, T.-C. Li, Y.-K. Tu, J.-C. Tsai, H.-S. Lai, and L.-T. Kuo, "Predicting long-term outcome after traumatic brain injury using repeated measurements of glasgow coma scale and data mining methods", *Journal of medical systems*, vol. 39, no. 2, p. 14, 2015.

[67] R. Lonkar, A. Dubrawski, M. Fiterau, and R. Garnett, "Mining intensive care vitals for leading indicators of adverse health events", *Emerg Health Threats J*, vol. 4, p. 11 073, 2011.

[68] M. Fiterau, A. Dubrawski, L. Chen, M. Hravnak, G Clermont, and M. R. Pinsky, "Automatic identification of artifacts in monitoring critically ill patients", *Intensive care medicine*, vol. 39, no. 2, S470, 2013.

[69] M. A. Pimentel, D. A. Clifton, and L. Tarassenko, "Gaussian process clustering for the functional characterisation of vital-sign trajectories", in *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, IEEE, 2013, pp. 1–6.

[70] M. Ferrario, F. Aletti, and G. Baselli, *The eu shockomics project international workshop at iccai'17*, 2017.

[71] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available: https://www.R-project.org/.

[72] S. Van Buuren, *Flexible imputation of missing data*. CRC press, 2012.

[73] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r", *Journal of Statistical Software*, vol. 45, no. 3, pp. 1–67, 2011. [Online]. Available: http://www.jstatsoft.org/v45/i03/.

[74] A. Kowarik and M. Templ, "Imputation with the R package VIM", *Journal of Statistical Software*, vol. 74, no. 7, pp. 1–16, 2016. DOI: 10.18637/jss.v074.i07.

[75] D. J. Stekhoven and P. Bühlmann, "Missforest—non-parametric missing value imputation for mixed-type data", *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2011.

[76] D. J. Stekhoven, "Missforest: Nonparametric missing value imputation using random forest", *Astrophysics Source Code Library*, 2015.

[77] J. L. Schafer, *Analysis of incomplete multivariate data*. CRC press, 1997.

[78] R. J. Little and D. B. Rubin, *Statistical analysis with missing data*. John Wiley & Sons, 2014, vol. 333.

[79] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[80] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers", *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[81] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels", in *Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.*, Ieee, 1999, pp. 41–48.

[82] O. Kramer, "K-nearest neighbors", in *Dimensionality Reduction with Unsupervised Nearest Neighbors*, Springer, 2013, pp. 13–23.

[83] T. F. Chan, G. H. Golub, and R. J. LeVeque, "Updating formulae and a pairwise algorithm for computing sample variances", in *COMPSTAT 1982 5th Symposium held at Toulouse 1982*, Springer, 1982, pp. 30–41.

[84] J. D. Rennie, L. Shih, J. Teevan, and D. R. Karger, "Tackling the poor assumptions of naive bayes text classifiers", in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 616–623.

[85] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics", *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[86] X. Lin, Q. Wang, P. Yin, L. Tang, Y. Tan, H. Li, K. Yan, and G. Xu, "A method for handling metabonomics data from liquid chromatography/mass spectrometry: Combinational use of support vector machine recursive feature elimination, genetic algorithm and random forest for feature selection", *Metabolomics*, vol. 7, no. 4, pp. 549–558, 2011.

[87] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing bayesian network structure learning algorithm", *Machine learning*, vol. 65, no. 1, pp. 31–78, 2006.

[88] N. Friedman *et al.*, "Learning belief networks in the presence of missing values and hidden variables", in *ICML*, vol. 97, 1997, pp. 125–133.

[89] N. Friedman, "The bayesian structural em algorithm", in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1998, pp. 129–138.

[90] C. Meek, "Graphical models: Selecting causal and statistical models", PhD thesis, PhD thesis, Carnegie Mellon University, 1997.

[91] A. E. Raftery, "Bayesian model selection in social research", *Sociological methodology*, pp. 111–163, 1995.

[92] M. Leone, P. Asfar, P. Radermacher, J.-L. Vincent, and C. Martin, "Optimizing mean arterial pressure in septic shock: A critical reappraisal of the literature", *Critical Care*, vol. 19, no. 1, p. 101, 2015.

Appendix A

Chapter 4: Complementary Tables and Figures

A.1 Dataset Overview

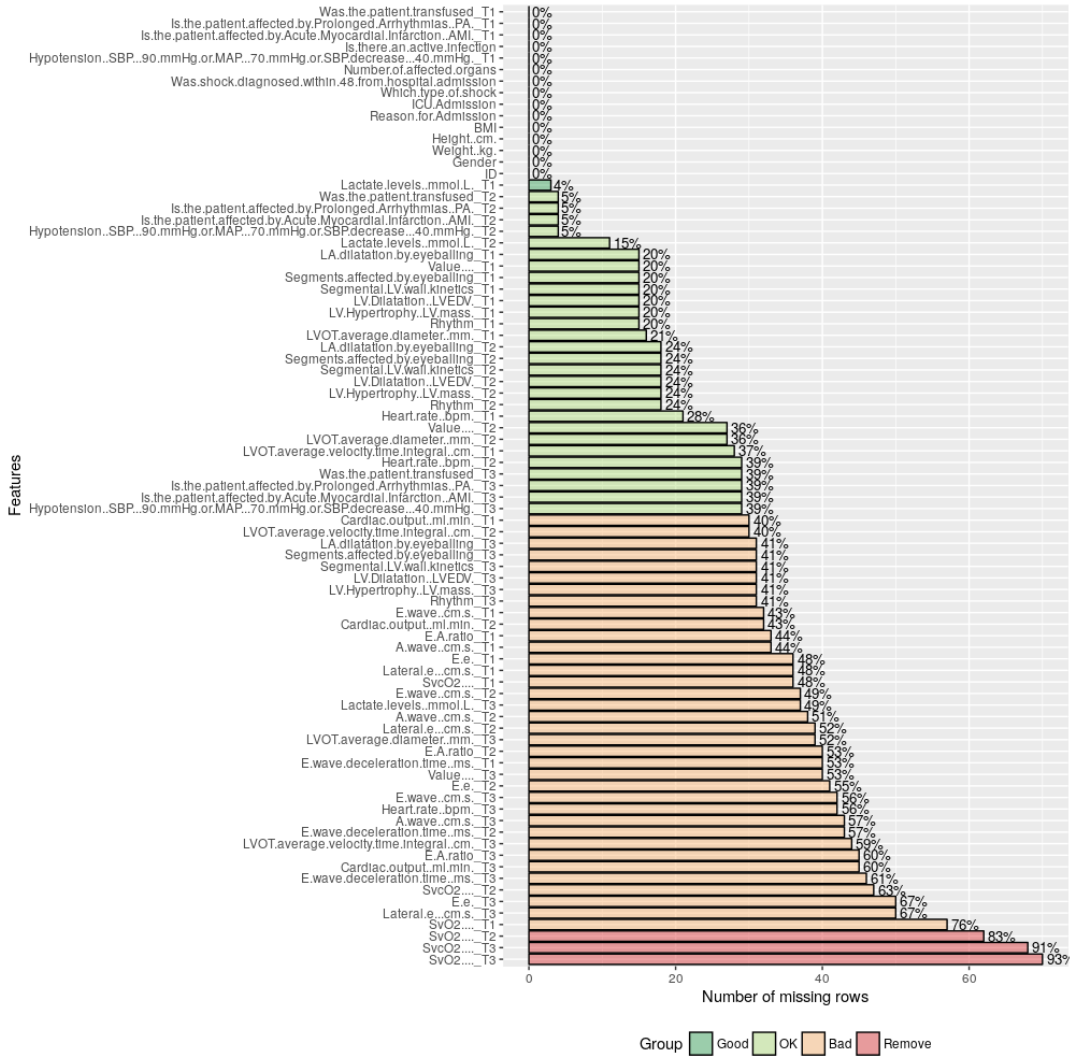


FIGURE A.1: The first 83 features of the ShockOmics data and their missing values.

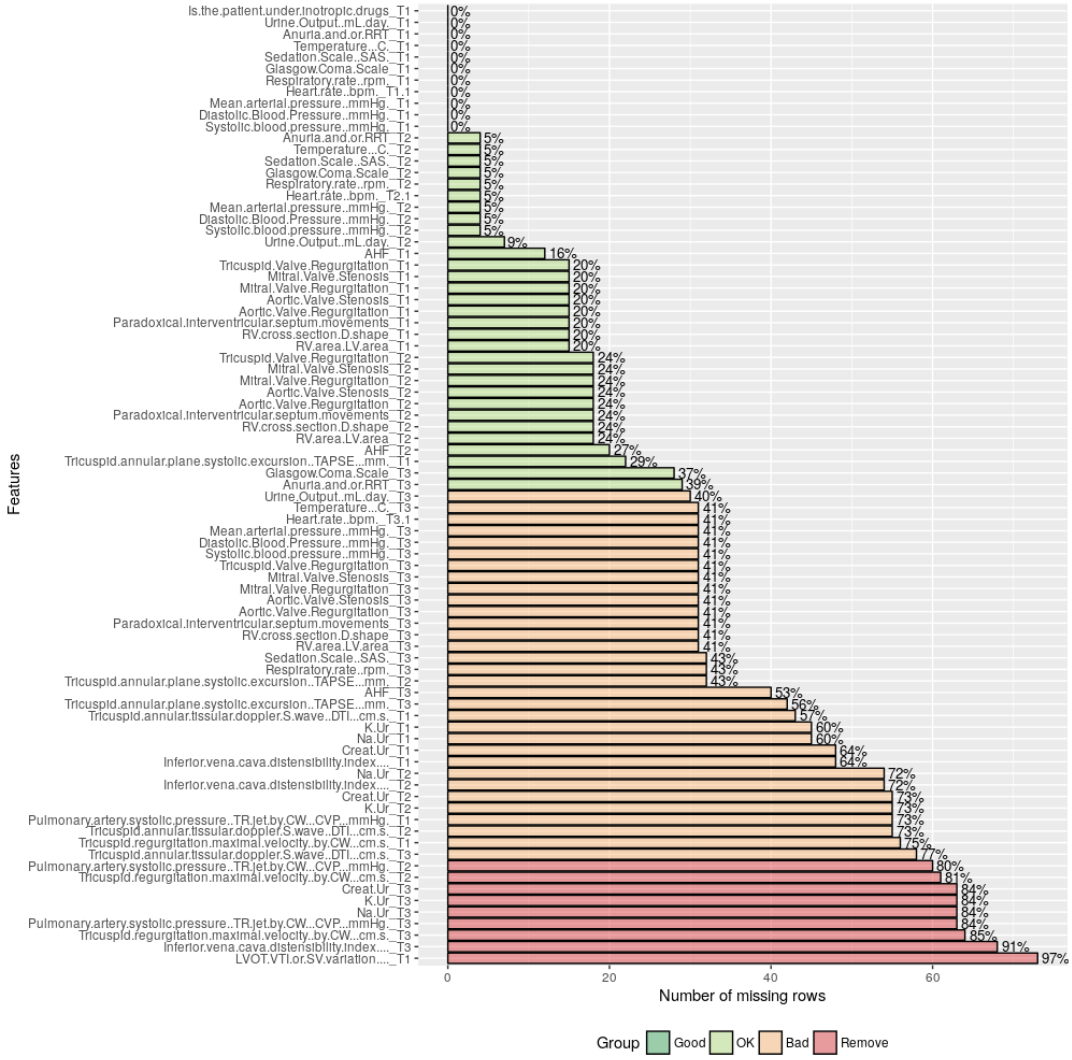


FIGURE A.2: Features of the ShockOmics data from 84 to 166 and their missing values.

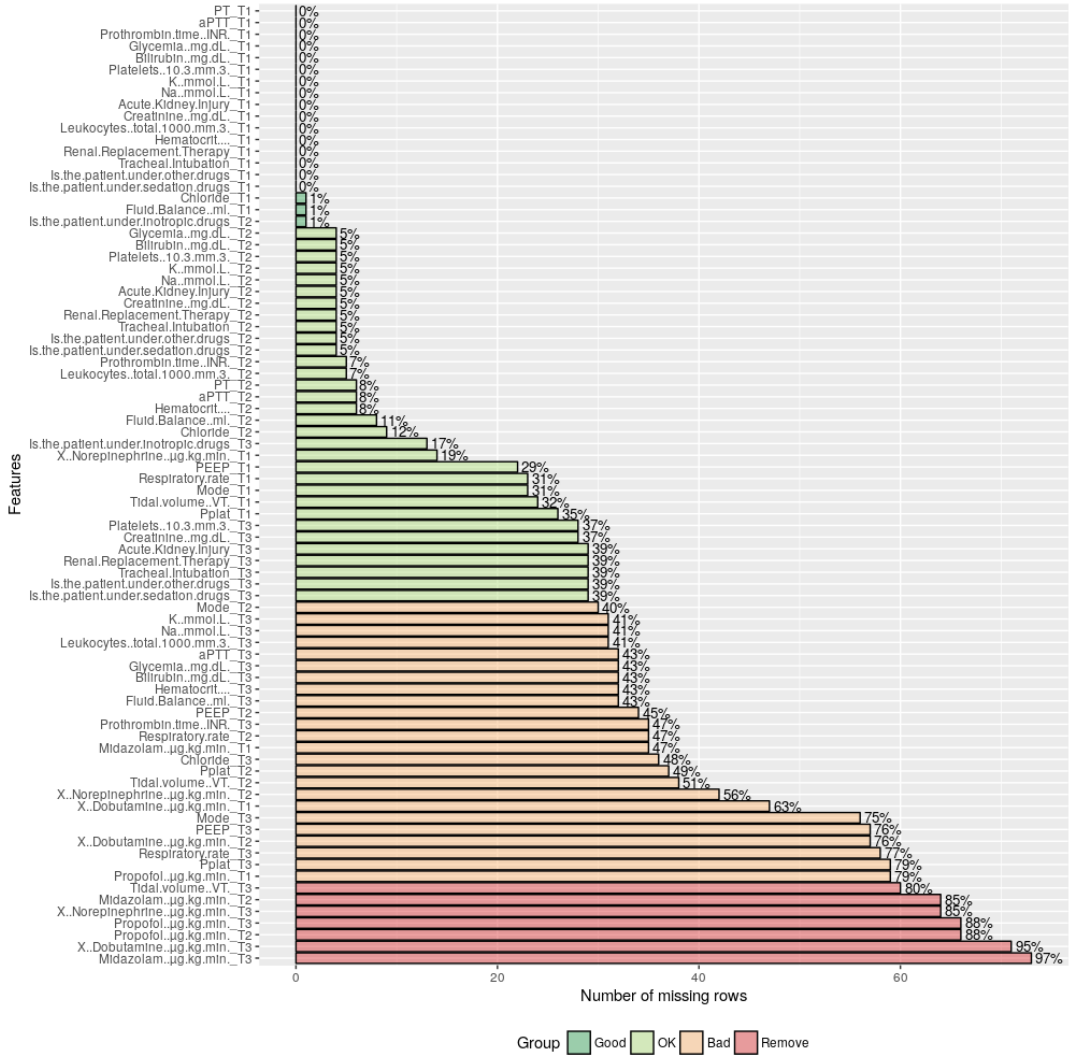


FIGURE A.3: Features of the ShockOmics data from 167 to 248 and their missing values.

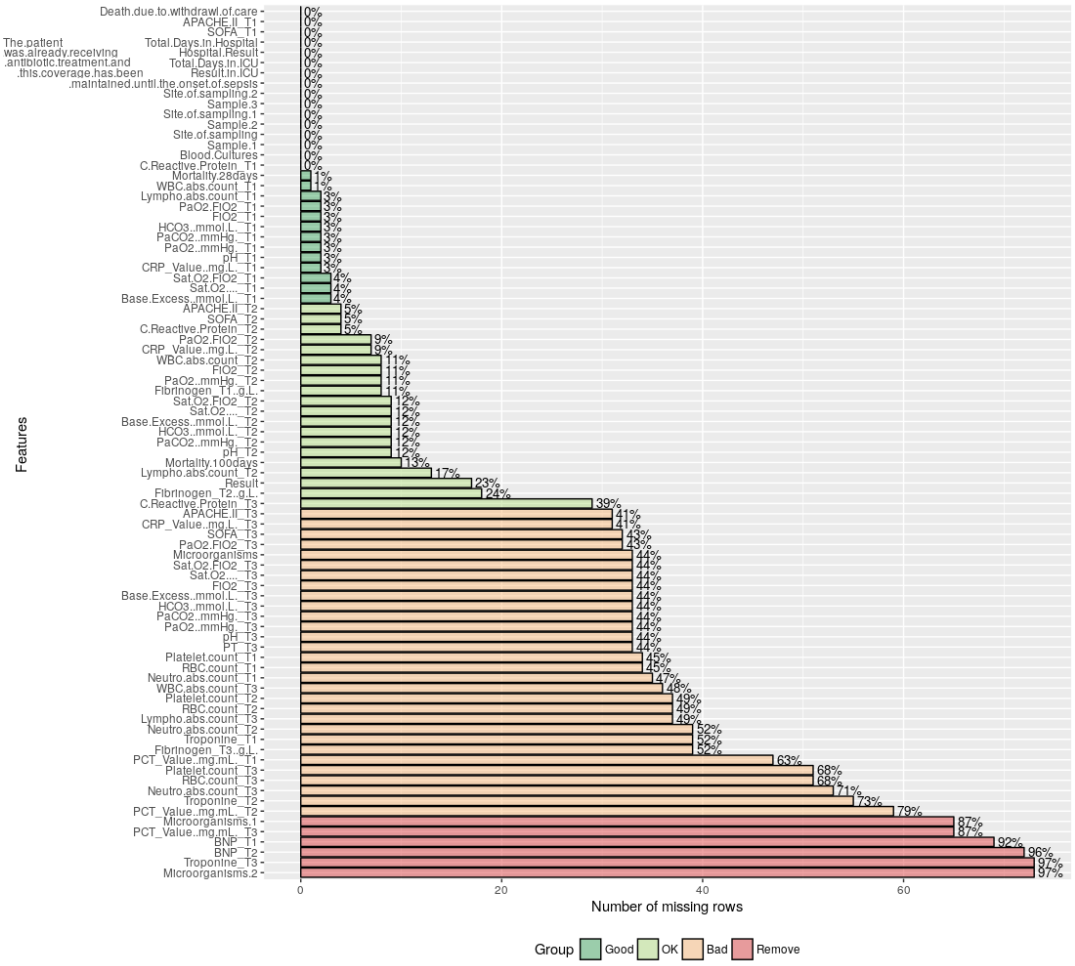


FIGURE A.4: Features of the ShockOmics data from 249 to 333 and their missing values.

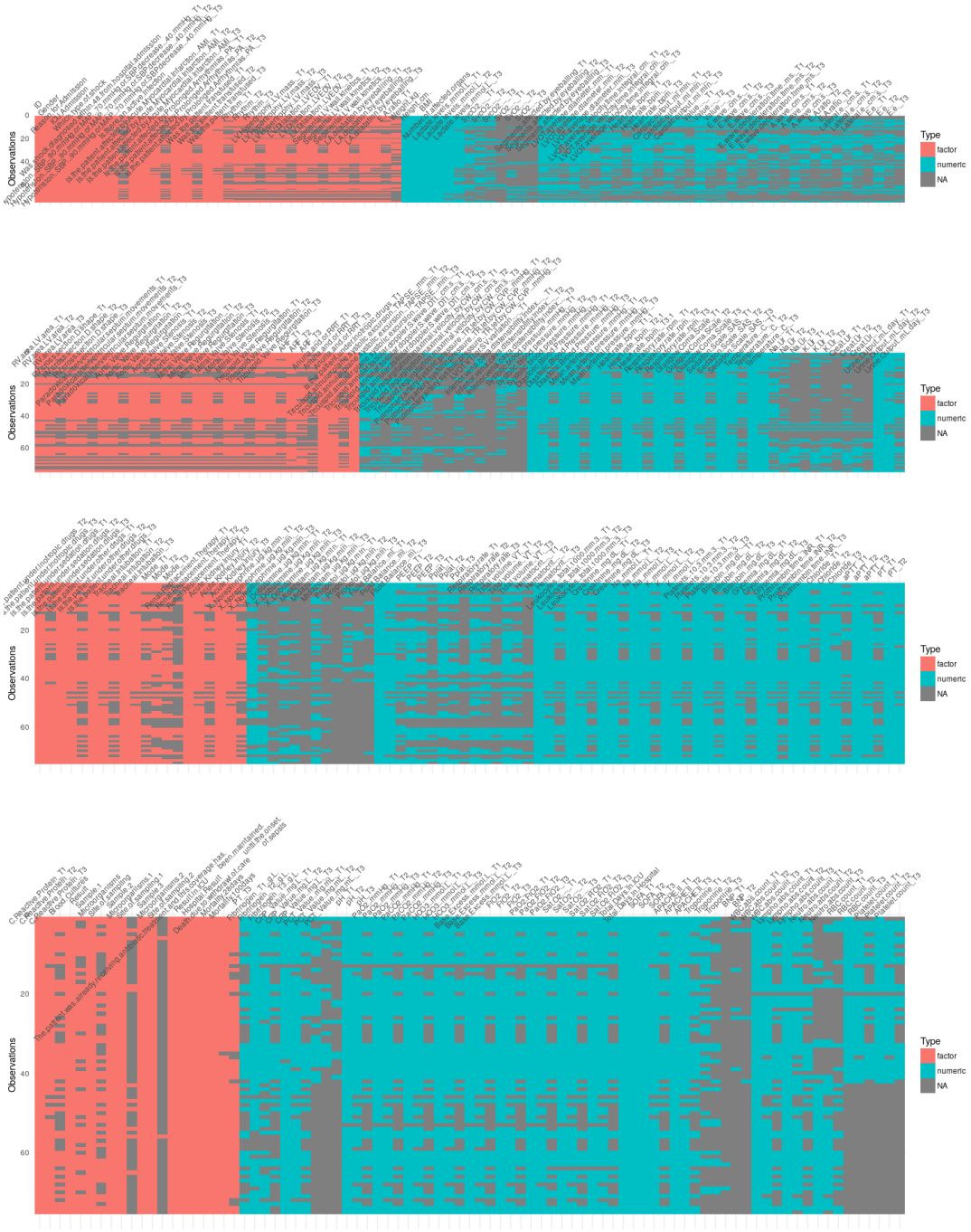


FIGURE A.5: Observations, features, their types and missing values of the ShockOmics dataset.

A.2 Bayesian Networks Configurations Performance

Struct.	Discr.	Accuracy	MCC
1	2	0.7368 ± 0.0666	0.2955 ± 0.1622
1	3	**0.8526 ± 0.0614**	0.6023 ± 0.1441
1	4	**0.8526 ± 0.1219**	**0.6381 ± 0.2973**
1	5	0.7158 ± 0.0855	0.0966 ± 0.1358
2	2	0.7263 ± 0.0614	0.2653 ± 0.138
2	3	0.8421 ± 0.0471	0.5723 ± 0.1157
2	4	0.8 ± 0.0614	0.1063 ± 0.1399
2	5	0.8316 ± 0.0211	0.0913 ± 0.1826
3	2	0.7579 ± 0.0537	0.2845 ± 0.1569
3	3	0.7895 ± 0.0333	0.364 ± 0.1988
3	4	0.8421 ± 0.0881	0.2749 ± 0.3367
3	5	0.7579 ± 0.0976	0.0 ± 0.0
4	2	0.6316 ± 0.0333	0.0357 ± 0.0662
4	3	0.7053 ± 0.0537	0.0585 ± 0.1893
4	4	0.7368 ± 0.0881	**0.1405 ± 0.2938**
4	5	0.6632 ± 0.1272	-0.1148 ± 0.0992
5	2	0.6842 ± 0.0333	0.0199 ± 0.1455
5	3	0.6316 ± 0.0577	0.0337 ± 0.1509
5	4	0.7053 ± 0.0976	0.0009 ± 0.1749
5	5	0.7684 ± 0.0537	-0.027 ± 0.0662
6	2	0.6632 ± 0.0714	0.1298 ± 0.1066
6	3	0.7263 ± 0.0394	0.0094 ± 0.1194
6	4	**0.7474 ± 0.0394**	-0.0092 ± 0.1216
6	5	0.7474 ± 0.1021	0.0982 ± 0.2392

TABLE A.1: Bayesian Networks structures, different number of variable discretization trained with the Maximum Likelihood Estimator: their accuracy and MCC. The best results in each column are highlighted in bold.

Struct.	Discr.	Sensitivity	Specificity	AUC
1	2	0.8336 ± 0.0989	0.4533 ± 0.2409	0.7368 ± 0.0666
1	3	0.9565 ± 0.0361	**0.5776 ± 0.1227**	**0.8526 ± 0.0614**
1	4	1.0 ± 0.0	0.5305 ± 0.3853	**0.8526 ± 0.1219**
1	5	0.9846 ± 0.0308	0.0667 ± 0.0816	0.7158 ± 0.0855
2	2	0.8167 ± 0.0834	0.4667 ± 0.1633	0.7263 ± 0.0614
2	3	0.959 ± 0.0538	0.5167 ± 0.1476	0.8421 ± 0.0471
2	4	**0.9857 ± 0.0286**	0.0733 ± 0.0904	0.8 ± 0.0614
2	5	1.0 ± 0.0	0.05 ± 0.1	0.8316 ± 0.0211
3	2	0.8758 ± 0.0514	0.3833 ± 0.1183	0.7579 ± 0.0537
3	3	1.0 ± 0.0	0.22 ± 0.1392	0.7895 ± 0.0333
3	4	1.0 ± 0.0	0.2 ± 0.2449	0.8421 ± 0.0881
3	5	1.0 ± 0.0	0.0 ± 0.0	0.7579 ± 0.0976
4	2	0.7971 ± 0.1142	0.25 ± 0.1667	0.6316 ± 0.0333
4	3	0.8215 ± 0.0814	0.2667 ± 0.2261	0.7053 ± 0.0537
4	4	**0.9571 ± 0.0857**	0.1167 ± 0.1453	0.7368 ± 0.0881
4	5	0.9261 ± 0.0607	0.0 ± 0.0	0.6632 ± 0.1272
5	2	0.9733 ± 0.0533	0.025 ± 0.05	0.6842 ± 0.0333
5	3	0.8328 ± 0.0519	0.2071 ± 0.117	0.6316 ± 0.0577
5	4	0.905 ± 0.0617	0.1067 ± 0.1373	0.7053 ± 0.0976
5	5	0.9464 ± 0.0527	0.04 ± 0.08	0.7684 ± 0.0537
6	2	0.8337 ± 0.1517	**0.3405 ± 0.3442**	0.6632 ± 0.0714
6	3	0.9082 ± 0.0873	0.09 ± 0.1114	0.7263 ± 0.0394
6	4	0.9317 ± 0.0631	0.0667 ± 0.1333	**0.7474 ± 0.0394**
6	5	0.9526 ± 0.0622	0.1071 ± 0.1317	0.7474 ± 0.1021

TABLE A.2: Bayesian Networks structures, different number of variable discretization trained with the Maximum Likelihood Estimator: their sensitivity, specificity and AUC. The best results in each column are highlighted in bold.

Struct.	Discr.	Accuracy	MCC
1	2	0.7789 ± 0.0842	0.3945 ± 0.2034
1	3	**0.8632 ± 0.0537**	**0.6181 ± 0.0854**
1	4	0.8421 ± 0.0577	0.5863 ± 0.1127
1	5	0.8 ± 0.0394	0.3313 ± 0.1707
2	2	0.7474 ± 0.0906	0.3471 ± 0.1915
2	3	0.8 ± 0.0774	0.4197 ± 0.1773
2	4	0.8421 ± 0.0744	0.368 ± 0.2782
2	5	0.7684 ± 0.0421	0.1607 ± 0.1998
3	2	0.8105 ± 0.0421	0.393 ± 0.1537
3	3	0.8526 ± 0.0211	0.1882 ± 0.2306
3	4	0.7158 ± 0.0855	0.153 ± 0.1931
3	5	0.7684 ± 0.0788	0.0 ± 0.0
4	2	0.6947 ± 0.0906	-0.0101 ± 0.1705
4	3	0.7053 ± 0.0632	0.0924 ± 0.168
4	4	0.7368 ± 0.0577	0.0502 ± 0.2463
4	5	0.7368 ± 0.1331	0.1104 ± 0.2311
5	2	0.7158 ± 0.1228	-0.1181 ± 0.1039
5	3	0.7158 ± 0.0421	0.0056 ± 0.1152
5	4	**0.7579 ± 0.0537**	0.1688 ± 0.2054
5	5	0.7263 ± 0.0516	-0.011 ± 0.0777
6	2	0.6947 ± 0.0394	-0.0143 ± 0.1935
6	3	0.7263 ± 0.0774	0.11 ± 0.1692
6	4	0.7474 ± 0.0842	**0.2249 ± 0.1893**
6	5	0.6947 ± 0.0906	-0.0034 ± 0.1997

TABLE A.3: Bayesian Networks structures, different number of variable discretization trained with the Bayesian Estimator: their accuracy and MCC. The best results in each column are highlighted in bold.

Struct.	Discr.	Sensitivity	Specificity	AUC
1	2	0.8897 ± 0.0658	0.4833 ± 0.2134	0.7789 ± 0.0842
1	3	0.9446 ± 0.0782	**0.5833 ± 0.1229**	**0.8632 ± 0.0537**
1	4	1.0 ± 0.0	0.4271 ± 0.1356	0.8421 ± 0.0577
1	5	0.9857 ± 0.0286	0.2133 ± 0.1306	0.8 ± 0.0394
2	2	0.8291 ± 0.0876	0.51 ± 0.1281	0.7474 ± 0.0906
2	3	0.9332 ± 0.0423	0.4319 ± 0.2023	0.8 ± 0.0774
2	4	0.9739 ± 0.0322	0.2733 ± 0.1665	0.8421 ± 0.0744
2	5	**1.0 ± 0.0**	0.0833 ± 0.1054	0.7684 ± 0.0421
3	2	0.8814 ± 0.0477	0.5033 ± 0.16	0.8105 ± 0.0421
3	3	0.9867 ± 0.0267	0.15 ± 0.2	0.8526 ± 0.0211
3	4	**1.0 ± 0.0**	0.0786 ± 0.102	0.7158 ± 0.0855
3	5	**1.0 ± 0.0**	0.0 ± 0.0	0.7684 ± 0.0788
4	2	0.9405 ± 0.0894	0.05 ± 0.1	0.6947 ± 0.0906
4	3	0.8213 ± 0.0824	0.24 ± 0.1679	0.7053 ± 0.0632
4	4	0.9198 ± 0.0658	0.115 ± 0.1513	0.7368 ± 0.0577
4	5	0.9509 ± 0.0406	0.1202 ± 0.1224	0.7368 ± 0.1331
5	2	0.9149 ± 0.0798	0.0 ± 0.0	0.7158 ± 0.1228
5	3	0.8692 ± 0.118	0.1667 ± 0.2108	0.7158 ± 0.0421
5	4	0.8814 ± 0.0654	0.2467 ± 0.1356	**0.7579 ± 0.0537**
5	5	0.9583 ± 0.0527	0.0286 ± 0.0571	0.7263 ± 0.0516
6	2	0.8687 ± 0.0549	0.1067 ± 0.1373	0.6947 ± 0.0394
6	3	0.8253 ± 0.0617	**0.2833 ± 0.1633**	0.7263 ± 0.0774
6	4	0.9414 ± 0.0541	0.2138 ± 0.1155	0.7474 ± 0.0842
6	5	**0.9542 ± 0.0651**	0.0333 ± 0.0667	0.6947 ± 0.0906

TABLE A.4: Bayesian Networks structures, different number of variable discretization trained with the Bayesian Estimator: their sensitivity, specificity and AUC. The best results in each column are highlighted in bold.

Appendix B

Chapter 5: Full+ dataset

B.1 Feature Selection Results

F	Accuracy	MCC	Sensitivity	Specificity	AUC
2	0.812 ± 0.072	0.503 ± 0.2	0.578 ± 0.194	0.896 ± 0.071	0.812 ± 0.072
3	0.846 ± 0.072	0.617 ± 0.183	0.716 ± 0.179	0.893 ± 0.08	0.846 ± 0.072
4	0.803 ± 0.066	0.502 ± 0.172	0.624 ± 0.199	0.867 ± 0.08	0.803 ± 0.066
5	0.807 ± 0.076	0.509 ± 0.196	0.618 ± 0.196	0.875 ± 0.084	0.807 ± 0.076
6	0.819 ± 0.076	0.544 ± 0.195	0.648 ± 0.202	0.88 ± 0.087	0.819 ± 0.076
7	0.843 ± 0.077	0.617 ± 0.193	0.744 ± 0.198	0.878 ± 0.081	0.843 ± 0.077
8	0.857 ± 0.074	0.654 ± 0.181	0.768 ± 0.18	0.889 ± 0.085	0.857 ± 0.074
9	0.862 ± 0.069	**0.667 ± 0.168**	**0.784 ± 0.178**	0.89 ± 0.076	0.862 ± 0.069
10	**0.863 ± 0.078**	0.663 ± 0.194	0.752 ± 0.184	0.903 ± 0.082	**0.863 ± 0.078**
11	0.838 ± 0.084	0.607 ± 0.208	0.734 ± 0.186	0.876 ± 0.087	0.838 ± 0.084
12	0.839 ± 0.067	0.62 ± 0.157	0.768 ± 0.189	0.864 ± 0.089	0.839 ± 0.067
13	0.851 ± 0.061	0.627 ± 0.151	0.704 ± 0.168	0.903 ± 0.079	0.851 ± 0.061
14	0.848 ± 0.066	0.609 ± 0.176	0.68 ± 0.179	**0.908 ± 0.066**	0.848 ± 0.066
15	0.835 ± 0.071	0.594 ± 0.175	0.706 ± 0.164	0.881 ± 0.082	0.835 ± 0.071
16	0.833 ± 0.077	0.577 ± 0.207	0.666 ± 0.212	0.893 ± 0.084	0.833 ± 0.077
17	0.829 ± 0.071	0.563 ± 0.192	0.662 ± 0.195	0.889 ± 0.073	0.829 ± 0.071
18	0.835 ± 0.068	0.584 ± 0.182	0.69 ± 0.193	0.887 ± 0.072	0.835 ± 0.068
19	0.826 ± 0.076	0.558 ± 0.201	0.654 ± 0.208	0.887 ± 0.082	0.826 ± 0.076
20	0.833 ± 0.074	0.581 ± 0.193	0.688 ± 0.199	0.884 ± 0.081	0.833 ± 0.074
21	0.843 ± 0.072	0.597 ± 0.204	0.696 ± 0.22	0.895 ± 0.072	0.843 ± 0.072
22	0.851 ± 0.081	0.624 ± 0.207	0.716 ± 0.198	0.899 ± 0.078	0.851 ± 0.081
23	0.831 ± 0.088	0.594 ± 0.199	0.724 ± 0.183	0.869 ± 0.102	0.831 ± 0.088
24	0.835 ± 0.08	0.59 ± 0.201	0.706 ± 0.184	0.881 ± 0.089	0.835 ± 0.08
25	0.843 ± 0.07	0.618 ± 0.167	0.736 ± 0.167	0.881 ± 0.082	0.843 ± 0.07
26	0.843 ± 0.077	0.622 ± 0.176	0.748 ± 0.166	0.876 ± 0.092	0.843 ± 0.077
27	0.844 ± 0.073	0.625 ± 0.166	0.748 ± 0.171	0.878 ± 0.091	0.844 ± 0.073
28	0.832 ± 0.077	0.583 ± 0.188	0.688 ± 0.184	0.884 ± 0.088	0.832 ± 0.077
29	0.828 ± 0.08	0.588 ± 0.189	0.718 ± 0.196	0.868 ± 0.103	0.828 ± 0.08
30	0.831 ± 0.088	0.585 ± 0.211	0.704 ± 0.201	0.876 ± 0.096	0.831 ± 0.088

TABLE B.1: The *UFS* features performance measures for **Full+** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
2	0.711 ± 0.053	-0.003 ± 0.166	0.046 ± 0.084	**0.948 ± 0.071**	0.711 ± 0.053
3	0.681 ± 0.073	0.038 ± 0.193	0.156 ± 0.122	0.868 ± 0.093	0.681 ± 0.073
4	0.733 ± 0.083	0.234 ± 0.254	0.312 ± 0.201	0.883 ± 0.092	0.733 ± 0.083
5	0.729 ± 0.079	0.213 ± 0.241	0.28 ± 0.172	0.889 ± 0.088	0.729 ± 0.079
6	0.74 ± 0.081	0.275 ± 0.229	0.352 ± 0.188	0.879 ± 0.092	0.74 ± 0.081
7	0.738 ± 0.101	0.339 ± 0.252	0.502 ± 0.211	0.823 ± 0.11	0.738 ± 0.101
8	0.719 ± 0.08	0.26 ± 0.2	0.414 ± 0.17	0.828 ± 0.098	0.719 ± 0.08
9	0.723 ± 0.089	0.27 ± 0.24	0.43 ± 0.214	0.827 ± 0.09	0.723 ± 0.089
10	0.692 ± 0.081	0.215 ± 0.207	0.428 ± 0.2	0.786 ± 0.103	0.692 ± 0.081
11	0.679 ± 0.088	0.162 ± 0.214	0.364 ± 0.191	0.791 ± 0.109	0.679 ± 0.088
12	0.719 ± 0.08	0.264 ± 0.218	0.436 ± 0.205	0.821 ± 0.1	0.719 ± 0.08
13	0.738 ± 0.095	0.35 ± 0.242	**0.54 ± 0.224**	0.809 ± 0.111	0.738 ± 0.095
14	0.713 ± 0.085	0.285 ± 0.212	0.494 ± 0.209	0.791 ± 0.105	0.713 ± 0.085
15	0.729 ± 0.099	0.329 ± 0.233	0.518 ± 0.225	0.805 ± 0.125	0.729 ± 0.099
16	**0.765 ± 0.083**	0.359 ± 0.237	0.428 ± 0.208	0.886 ± 0.097	**0.765 ± 0.083**
17	0.751 ± 0.064	0.316 ± 0.2	0.41 ± 0.216	0.872 ± 0.093	0.751 ± 0.064
18	0.748 ± 0.078	0.288 ± 0.246	0.354 ± 0.194	0.889 ± 0.086	0.748 ± 0.078
19	0.739 ± 0.076	0.284 ± 0.214	0.388 ± 0.189	0.864 ± 0.085	0.739 ± 0.076
20	0.763 ± 0.076	**0.354 ± 0.21**	0.422 ± 0.206	0.884 ± 0.1	0.763 ± 0.076
21	0.739 ± 0.084	0.284 ± 0.235	0.396 ± 0.214	0.861 ± 0.1	0.739 ± 0.084
22	0.747 ± 0.079	0.313 ± 0.227	0.416 ± 0.199	0.865 ± 0.091	0.747 ± 0.079
23	0.739 ± 0.082	0.289 ± 0.232	0.398 ± 0.218	0.861 ± 0.095	0.739 ± 0.082
24	0.741 ± 0.082	0.287 ± 0.232	0.388 ± 0.215	0.866 ± 0.103	0.741 ± 0.082
25	0.735 ± 0.076	0.269 ± 0.22	0.376 ± 0.21	0.863 ± 0.095	0.735 ± 0.076
26	0.724 ± 0.071	0.244 ± 0.194	0.362 ± 0.185	0.854 ± 0.098	0.724 ± 0.071
27	0.725 ± 0.08	0.256 ± 0.222	0.392 ± 0.206	0.844 ± 0.098	0.725 ± 0.08
28	0.722 ± 0.09	0.244 ± 0.239	0.368 ± 0.191	0.849 ± 0.1	0.722 ± 0.09
29	0.748 ± 0.088	0.312 ± 0.256	0.414 ± 0.235	0.867 ± 0.103	0.748 ± 0.088
30	0.739 ± 0.085	0.284 ± 0.244	0.39 ± 0.222	0.864 ± 0.094	0.739 ± 0.085

TABLE B.2: The *RFE* features performance measures for the **Full+** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
17	0.748 ± 0.081	0.31 ± 0.233	0.388 ± 0.196	0.877 ± 0.1	0.748 ± 0.081
19	0.719 ± 0.084	0.228 ± 0.235	0.354 ± 0.213	0.849 ± 0.107	0.719 ± 0.084

TABLE B.3: The *RFECV* features performance measures for the **Full+** dataset. The first entry with 17 features used the *average precision* metric. The second one used the *accuracy*.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
2	0.802 ± 0.09	0.491 ± 0.198	0.576 ± 0.168	0.882 ± 0.112	0.802 ± 0.09
3	0.818 ± 0.111	0.54 ± 0.25	0.604 ± 0.21	0.894 ± 0.125	0.818 ± 0.111
4	**0.842 ± 0.104**	0.6 ± 0.235	0.664 ± 0.196	**0.905 ± 0.11**	**0.842 ± 0.104**
5	0.793 ± 0.114	0.51 ± 0.232	0.66 ± 0.175	0.841 ± 0.14	0.793 ± 0.114
6	0.8 ± 0.098	0.525 ± 0.209	0.696 ± 0.178	0.837 ± 0.118	0.8 ± 0.098
7	0.782 ± 0.109	0.496 ± 0.217	**0.698 ± 0.187**	0.812 ± 0.129	0.782 ± 0.109
8	0.794 ± 0.09	0.504 ± 0.216	0.674 ± 0.191	0.836 ± 0.101	0.794 ± 0.09
9	0.814 ± 0.079	0.537 ± 0.187	0.642 ± 0.184	0.876 ± 0.096	0.814 ± 0.079
10	0.769 ± 0.106	0.439 ± 0.219	0.602 ± 0.182	0.829 ± 0.131	0.769 ± 0.106
11	0.806 ± 0.083	0.517 ± 0.191	0.63 ± 0.182	0.869 ± 0.101	0.806 ± 0.083
12	0.787 ± 0.105	0.47 ± 0.233	0.584 ± 0.205	0.859 ± 0.128	0.787 ± 0.105
13	0.788 ± 0.109	0.482 ± 0.227	0.61 ± 0.188	0.851 ± 0.135	0.788 ± 0.109
14	0.813 ± 0.097	0.541 ± 0.208	0.654 ± 0.198	0.87 ± 0.121	0.813 ± 0.097
15	0.818 ± 0.095	0.558 ± 0.206	0.676 ± 0.167	0.869 ± 0.114	0.818 ± 0.095
16	0.78 ± 0.096	0.45 ± 0.234	0.592 ± 0.217	0.847 ± 0.117	0.78 ± 0.096
17	0.789 ± 0.112	0.48 ± 0.242	0.604 ± 0.214	0.855 ± 0.142	0.789 ± 0.112
18	0.809 ± 0.083	0.515 ± 0.196	0.618 ± 0.198	0.877 ± 0.1	0.809 ± 0.083
19	0.828 ± 0.075	0.559 ± 0.197	0.632 ± 0.203	0.898 ± 0.088	0.828 ± 0.075
20	0.825 ± 0.086	0.555 ± 0.216	0.62 ± 0.193	0.899 ± 0.096	0.825 ± 0.086
21	0.81 ± 0.093	0.53 ± 0.212	0.638 ± 0.189	0.871 ± 0.11	0.81 ± 0.093
22	0.808 ± 0.086	0.521 ± 0.213	0.64 ± 0.21	0.868 ± 0.103	0.808 ± 0.086
23	0.833 ± 0.083	0.583 ± 0.193	0.684 ± 0.165	0.886 ± 0.093	0.833 ± 0.083
24	0.816 ± 0.072	0.535 ± 0.171	0.626 ± 0.162	0.884 ± 0.088	0.816 ± 0.072
25	0.815 ± 0.087	0.539 ± 0.209	0.646 ± 0.194	0.876 ± 0.099	0.815 ± 0.087
26	0.827 ± 0.086	0.568 ± 0.216	0.676 ± 0.205	0.881 ± 0.093	0.827 ± 0.086
27	0.831 ± 0.081	0.583 ± 0.193	0.68 ± 0.192	0.885 ± 0.097	0.831 ± 0.081
28	0.821 ± 0.085	0.553 ± 0.218	0.674 ± 0.203	0.874 ± 0.09	0.821 ± 0.085
29	0.823 ± 0.086	0.558 ± 0.205	0.66 ± 0.189	0.881 ± 0.096	0.823 ± 0.086
30	0.841 ± 0.079	**0.605 ± 0.189**	0.688 ± 0.182	0.896 ± 0.094	0.841 ± 0.079

TABLE B.4: The *UFS+RFE* features performance measures for the **Full+** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
2	0.839 ± 0.064	0.571 ± 0.18	0.55 ± 0.195	**0.943 ± 0.071**	0.839 ± 0.064
3	0.822 ± 0.08	0.532 ± 0.214	0.58 ± 0.189	0.908 ± 0.084	0.822 ± 0.08
4	0.844 ± 0.067	0.581 ± 0.194	0.6 ± 0.194	0.931 ± 0.06	0.844 ± 0.067
5	0.848 ± 0.067	0.615 ± 0.171	0.686 ± 0.175	0.906 ± 0.075	0.848 ± 0.067
6	0.807 ± 0.077	0.517 ± 0.199	0.64 ± 0.208	0.867 ± 0.087	0.807 ± 0.077
7	0.831 ± 0.07	0.57 ± 0.186	0.66 ± 0.184	0.891 ± 0.078	0.831 ± 0.07
8	0.839 ± 0.081	0.613 ± 0.201	**0.754 ± 0.194**	0.87 ± 0.088	0.839 ± 0.081
9	0.818 ± 0.073	0.556 ± 0.178	0.682 ± 0.192	0.867 ± 0.095	0.818 ± 0.073
10	0.808 ± 0.066	0.515 ± 0.172	0.632 ± 0.178	0.871 ± 0.076	0.808 ± 0.066
11	0.799 ± 0.072	0.48 ± 0.191	0.58 ± 0.184	0.877 ± 0.08	0.799 ± 0.072
12	0.785 ± 0.076	0.457 ± 0.203	0.584 ± 0.215	0.857 ± 0.093	0.785 ± 0.076
13	0.808 ± 0.081	0.506 ± 0.216	0.612 ± 0.219	0.878 ± 0.085	0.808 ± 0.081
14	0.821 ± 0.086	0.568 ± 0.213	0.722 ± 0.2	0.856 ± 0.092	0.821 ± 0.086
15	0.824 ± 0.077	0.549 ± 0.204	0.634 ± 0.212	0.892 ± 0.085	0.824 ± 0.077
16	0.844 ± 0.08	0.623 ± 0.184	0.716 ± 0.186	0.889 ± 0.104	0.844 ± 0.08
17	0.836 ± 0.08	0.597 ± 0.198	0.71 ± 0.201	0.881 ± 0.094	0.836 ± 0.08
18	0.83 ± 0.08	0.571 ± 0.207	0.672 ± 0.201	0.886 ± 0.085	0.83 ± 0.08
19	0.843 ± 0.079	0.608 ± 0.208	0.708 ± 0.212	0.891 ± 0.09	0.843 ± 0.079
20	0.838 ± 0.077	0.605 ± 0.183	0.722 ± 0.194	0.879 ± 0.092	0.838 ± 0.077
21	0.855 ± 0.068	0.638 ± 0.183	0.724 ± 0.193	0.901 ± 0.08	0.855 ± 0.068
22	0.845 ± 0.071	0.605 ± 0.188	0.684 ± 0.196	0.903 ± 0.074	0.845 ± 0.071
23	0.841 ± 0.076	0.606 ± 0.192	0.702 ± 0.197	0.89 ± 0.089	0.841 ± 0.076
24	0.849 ± 0.064	0.63 ± 0.166	0.74 ± 0.193	0.888 ± 0.082	0.849 ± 0.064
25	0.848 ± 0.062	0.622 ± 0.171	0.738 ± 0.191	0.887 ± 0.069	0.848 ± 0.062
26	0.859 ± 0.064	0.649 ± 0.172	0.744 ± 0.196	0.9 ± 0.07	0.859 ± 0.064
27	0.854 ± 0.062	0.63 ± 0.172	0.708 ± 0.205	0.906 ± 0.071	0.854 ± 0.062
28	**0.861 ± 0.062**	**0.655 ± 0.161**	0.748 ± 0.169	0.901 ± 0.069	**0.861 ± 0.062**
29	0.843 ± 0.074	0.615 ± 0.193	0.72 ± 0.212	0.886 ± 0.094	0.843 ± 0.074
30	0.852 ± 0.07	0.628 ± 0.193	0.706 ± 0.216	0.904 ± 0.084	0.852 ± 0.07

TABLE B.5: The *RF* performance measures for the **Full+** dataset.

B.2 Feature Sets Rankings

R	Univariate feature selection	Recursive feature elimination
1	Inferior vena cava distensibility index _T3	Hematocrit _T1
2	SOFA_T2	K Ur_T1
3	APACHE II_T1	Tricuspid regurgitation max vel _T2
4	APACHE II_T2	Platelet count_T1
5	APACHE II_T3	Creat Ur_T1
6	SOFA_T3	PT_T1
7	SOFA_T1	Tricuspid regurgitation max vel _T1
8	LA dilatation by eyeballing_T1=Yes	Respiratory rate rpm _T1
9	LA dilatation by eyeballing_T1=No	Tidal volume VT _T3
10	X Norepinephrine xb5g kg min _T2	Tidal volume VT _T1
11	Glasgow Coma Scale_T2	aPTT_T1
12	FiO2_T3	HCO3 mmol L _T1
13	E e _T3	E wave cm s _T1
14	pH_T2	Creat Ur_T3
15	Lactate levels mmol L _T2	Diastolic Blood Pressure mmHg _T3
16	PCT_Value mg mL _T2	Neutro abs count_T2
17	Sedation Scale SAS _T2	Value _T1
18	Na Ur_T3	Heart rate bpm _T2 1
19	FiO2_T2	K Ur_T3
20	Respiratory rate_T1	APACHE II_T2
21	Tricuspid regurgitation max vel _T1	Leukocytes total 1000 mm 3 _T2
22	Tracheal Intubation_T2=Yes	PCT_Value mg mL _T2
23	Tracheal Intubation_T2=No	CRP_Value mg L _T3
24	Sat O2 FiO2_T3	WBC abs count_T2
25	X Norepinephrine xb5g kg min _T1	SvO2 _T3
26	Creatinine mg dL _T3	Mean arterial pressure mmHg _T3
27	Sat O2 FiO2_T2	Leukocytes total 1000 mm 3 _T3
28	Hypotension,MAP or SBP decrease_T2=Yes	Inferior vena cava distensibility index _T1
29	Hypotension,MAP or SBP decrease_T2=No	APACHE II_T1
30	Number of affected organs	PaCO2 mmHg _T1

TABLE B.6: The *UFS* and the *RFE* feature rankings for the **Full+** dataset.

```
AVERAGE PRECISION: Creat Ur_T1, Creat Ur_T3, Diastolic Blood Pressure
    mmHg_T3, E wave cm s _T1, HCO3 mmol L _T1, Hematocrit _T1, K Ur_T1,
     Neutro abs count_T2, PT_T1, Platelet count_T1, Respiratory rate
    rpm _T1, Tidal volume VT _T1, Tidal volume VT _T3, Tricuspid
    regurgitation maximal velocity by CW cm s _T1, Tricuspid
    regurgitation maximal velocity by CW cm s _T2, Value _T1, aPTT_T1;

ACCURACY: Creat Ur_T1, Creat Ur_T3, Diastolic Blood Pressure mmHg _T3,
    E wave cm s _T1, HCO3 mmol L _T1, Heart rate bpm _T2 1, Hematocrit
    _T1, K Ur_T1, K Ur_T3, Neutro abs count_T2, PT_T1, Platelet
    count_T1, Respiratory rate rpm _T1, Tidal volume VT _T1, Tidal
    volume VT _T3, Tricuspid regurgitation maximal velocity by CW cm s
    _T1, Tricuspid regurgitation maximal velocity by CW cm s _T2, Value
     _T1, aPTT_T1;
```

List of features for the *RFECV* with the *average precision* and the *accuracy* metrics for the **Full+** dataset.

R	UFS+FSE	Random Forest
1	Inferior vena cava distensibility index_T3	Inferior vena cava distensibility index_T3
2	Tracheal Intubation_T2=No	Lactate levels mmol L _T2
3	X Dobutamine xb5g kg min_T3	SOFA_T2
4	LVOT average diameter mm _T1	SOFA_T3
5	Tracheal Intubation_T2=Yes	Na Ur_T3
6	Glasgow Coma Scale_T3	APACHE II_T2
7	Glasgow Coma Scale_T1	FiO2_T3
8	PCT_Value mg mL _T3	APACHE II_T1
9	K Ur_T1	PCT_Value_T1
10	PCT_Value mg mL _T1	X Norepinephrine _T2
11	APACHE II_T3	K Ur_T2
12	Heart rate bpm _T3 1	APACHE II_T3
13	Lactate levels mmol L _T1	Urine Output_T1
14	Tricuspid regurgitation max vel_T1	Respiratory rate_T1
15	SOFA_T1	Sat O2 FiO2_T3
16	Respiratory rate_T1	Pplat_T3
17	PT_T2	X Dobutamine_T3
18	Platelet count_T1	Lateral_T3
19	Platelets _T3	Tricuspid regurgitation max vel_T3
20	LA dilatation by eyeballing_T1=No	Tidal volume VT _T2
21	LA dilatation by eyeballing_T1=Yes	Heart rate bpm _T3 1
22	Glasgow Coma Scale_T2	E e _T3
23	Number of affected organs	E wave_T1
24	APACHE II_T2	X Norepinephrine_T1
25	APACHE II_T1	Tricuspid annular plane syst excursion_T3
26	SOFA_T3	Na Ur_T2
27	CRP_Value mg L _T3	CRP_Value_T3
28	Na Ur_T3	Urine Output _T2
29	Platelet count_T2	Glasgow Coma Scale_T3
30	Tidal volume VT_T1	SOFA_T1

TABLE B.7: The *UFS+RFE* and the *RF* feature rankings for the **Full+** dataset.

Appendix C

Chapter 5: Full dataset

C.1 Feature Selection Results

F	Accuracy	MCC	Sensitivity	Specificity	AUC
2	0.848 ± 0.071	0.615 ± 0.19	0.71 ± 0.205	0.897 ± 0.072	0.848 ± 0.071
3	0.802 ± 0.072	0.496 ± 0.196	0.62 ± 0.203	0.866 ± 0.082	0.802 ± 0.072
4	0.82 ± 0.082	0.572 ± 0.19	0.732 ± 0.173	0.851 ± 0.09	0.82 ± 0.082
5	0.832 ± 0.08	0.586 ± 0.201	0.712 ± 0.197	0.874 ± 0.088	0.832 ± 0.08
6	0.814 ± 0.074	0.546 ± 0.173	0.688 ± 0.18	0.859 ± 0.091	0.814 ± 0.074
7	0.803 ± 0.084	0.517 ± 0.215	0.68 ± 0.214	0.847 ± 0.086	0.803 ± 0.084
8	0.801 ± 0.081	0.494 ± 0.219	0.622 ± 0.208	0.865 ± 0.084	0.801 ± 0.081
9	0.795 ± 0.083	0.471 ± 0.22	0.578 ± 0.217	0.873 ± 0.092	0.795 ± 0.083
10	0.824 ± 0.088	0.553 ± 0.224	0.654 ± 0.217	0.884 ± 0.09	0.824 ± 0.088
11	0.823 ± 0.077	0.543 ± 0.212	0.636 ± 0.199	0.889 ± 0.081	0.823 ± 0.077
12	0.811 ± 0.097	0.539 ± 0.218	0.668 ± 0.198	0.862 ± 0.114	0.811 ± 0.097
13	0.802 ± 0.096	0.543 ± 0.187	0.726 ± 0.149	0.829 ± 0.119	0.802 ± 0.096
14	0.796 ± 0.11	0.534 ± 0.23	0.722 ± 0.187	0.823 ± 0.131	0.796 ± 0.11
15	0.832 ± 0.096	0.599 ± 0.213	0.734 ± 0.186	0.867 ± 0.11	0.832 ± 0.096
16	0.843 ± 0.097	0.64 ± 0.194	0.79 ± 0.158	0.862 ± 0.115	0.843 ± 0.097
17	0.825 ± 0.106	0.595 ± 0.232	0.76 ± 0.185	0.849 ± 0.116	0.825 ± 0.106
18	0.836 ± 0.097	0.616 ± 0.209	0.764 ± 0.163	0.861 ± 0.108	0.836 ± 0.097
19	0.829 ± 0.085	0.614 ± 0.181	**0.798 ± 0.166**	0.84 ± 0.103	0.829 ± 0.085
20	0.853 ± 0.085	0.653 ± 0.201	0.79 ± 0.184	0.876 ± 0.097	0.853 ± 0.085
21	0.845 ± 0.081	0.635 ± 0.186	0.776 ± 0.161	0.87 ± 0.09	0.845 ± 0.081
22	0.845 ± 0.095	0.635 ± 0.207	0.768 ± 0.159	0.872 ± 0.106	0.845 ± 0.095
23	0.85 ± 0.081	0.643 ± 0.18	0.764 ± 0.153	0.881 ± 0.096	0.85 ± 0.081
24	**0.872 ± 0.073**	**0.687 ± 0.173**	0.782 ± 0.147	**0.904 ± 0.079**	**0.872 ± 0.073**
25	0.841 ± 0.089	0.621 ± 0.207	0.752 ± 0.175	0.873 ± 0.097	0.841 ± 0.089
26	0.832 ± 0.083	0.597 ± 0.199	0.754 ± 0.189	0.859 ± 0.084	0.832 ± 0.083
27	0.832 ± 0.079	0.602 ± 0.189	0.746 ± 0.192	0.862 ± 0.096	0.832 ± 0.079
28	0.832 ± 0.062	0.59 ± 0.156	0.728 ± 0.176	0.869 ± 0.079	0.832 ± 0.062
29	0.827 ± 0.069	0.59 ± 0.161	0.75 ± 0.163	0.854 ± 0.086	0.827 ± 0.069
30	0.851 ± 0.068	0.636 ± 0.17	0.756 ± 0.191	0.884 ± 0.077	0.851 ± 0.068

TABLE C.1: The *UFS* features performance measures for the **Full** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
2	0.72 ± 0.077	0.223 ± 0.232	0.348 ± 0.213	0.853 ± 0.098	0.72 ± 0.077
3	0.755 ± 0.069	0.335 ± 0.208	0.436 ± 0.22	0.869 ± 0.087	0.755 ± 0.069
4	0.759 ± 0.083	0.437 ± 0.204	0.662 ± 0.217	0.794 ± 0.098	0.759 ± 0.083
5	0.767 ± 0.092	0.466 ± 0.224	0.7 ± 0.225	0.791 ± 0.104	0.767 ± 0.092
6	0.783 ± 0.081	0.508 ± 0.196	$\mathbf{0.73 \pm 0.218}$	0.802 ± 0.101	0.783 ± 0.081
7	0.76 ± 0.092	0.435 ± 0.214	0.638 ± 0.226	0.804 ± 0.121	0.76 ± 0.092
8	0.77 ± 0.096	0.456 ± 0.223	0.654 ± 0.222	0.811 ± 0.121	0.77 ± 0.096
9	0.78 ± 0.08	0.482 ± 0.181	0.66 ± 0.195	0.823 ± 0.112	0.78 ± 0.08
10	0.806 ± 0.085	0.524 ± 0.213	0.662 ± 0.226	0.857 ± 0.105	0.806 ± 0.085
11	0.822 ± 0.088	$\mathbf{0.559 \pm 0.222}$	0.682 ± 0.219	0.871 ± 0.096	0.822 ± 0.088
12	0.803 ± 0.075	0.51 ± 0.188	0.644 ± 0.199	0.86 ± 0.097	0.803 ± 0.075
13	$\mathbf{0.828 \pm 0.083}$	0.546 ± 0.229	0.592 ± 0.242	0.912 ± 0.084	$\mathbf{0.828 \pm 0.083}$
14	0.825 ± 0.087	0.547 ± 0.22	0.6 ± 0.204	0.905 ± 0.095	0.825 ± 0.087
15	0.806 ± 0.082	0.484 ± 0.233	0.548 ± 0.22	0.899 ± 0.082	0.806 ± 0.082
16	0.751 ± 0.066	0.221 ± 0.271	0.256 ± 0.252	$\mathbf{0.927 \pm 0.08}$	0.751 ± 0.066
17	0.759 ± 0.065	0.265 ± 0.26	0.292 ± 0.242	0.926 ± 0.072	0.759 ± 0.065
18	0.75 ± 0.075	0.227 ± 0.269	0.254 ± 0.228	$\mathbf{0.927 \pm 0.087}$	0.75 ± 0.075
19	0.759 ± 0.066	0.266 ± 0.255	0.296 ± 0.239	0.925 ± 0.074	0.759 ± 0.066
20	0.783 ± 0.071	0.373 ± 0.24	0.38 ± 0.227	0.926 ± 0.067	0.783 ± 0.071
21	0.774 ± 0.069	0.335 ± 0.247	0.334 ± 0.232	0.931 ± 0.076	0.774 ± 0.069
22	0.769 ± 0.067	0.308 ± 0.246	0.316 ± 0.225	0.931 ± 0.071	0.769 ± 0.067
23	0.77 ± 0.063	0.321 ± 0.238	0.332 ± 0.232	0.926 ± 0.072	0.77 ± 0.063
24	0.765 ± 0.075	0.299 ± 0.268	0.336 ± 0.241	0.918 ± 0.08	0.765 ± 0.075
25	0.757 ± 0.065	0.287 ± 0.235	0.336 ± 0.233	0.907 ± 0.077	0.757 ± 0.065
26	0.765 ± 0.073	0.317 ± 0.253	0.348 ± 0.243	0.914 ± 0.086	0.765 ± 0.073
27	0.777 ± 0.07	0.381 ± 0.223	0.424 ± 0.242	0.904 ± 0.092	0.777 ± 0.07
28	0.774 ± 0.065	0.36 ± 0.234	0.412 ± 0.25	0.903 ± 0.084	0.774 ± 0.065
29	0.778 ± 0.067	0.389 ± 0.204	0.442 ± 0.21	0.898 ± 0.086	0.778 ± 0.067
30	0.778 ± 0.072	0.397 ± 0.216	0.47 ± 0.23	0.888 ± 0.087	0.778 ± 0.072

TABLE C.2: The *RFE* features performance measures for the **Full** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
12	0.788 ± 0.088	0.468 ± 0.223	0.602 ± 0.213	0.854 ± 0.107	0.788 ± 0.088
23	0.756 ± 0.071	0.292 ± 0.247	0.338 ± 0.233	0.906 ± 0.087	0.756 ± 0.071

TABLE C.3: The *RFECV* features performance measures for the **Full** dataset. The first entry with 12 features used the *precision* metric. The second one used the *AUC*.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
2	0.767 ± 0.06	0.249 ± 0.208	0.144 ± 0.147	0.99 ± 0.099	0.767 ± 0.06
3	0.771 ± 0.027	0.248 ± 0.196	0.128 ± 0.104	**1.0 ± 0.0**	0.771 ± 0.027
4	0.776 ± 0.032	0.276 ± 0.205	0.15 ± 0.121	**1.0 ± 0.0**	0.776 ± 0.032
5	0.775 ± 0.031	0.28 ± 0.195	0.15 ± 0.111	0.999 ± 0.014	0.775 ± 0.031
6	0.773 ± 0.03	0.262 ± 0.201	0.142 ± 0.118	0.999 ± 0.014	0.773 ± 0.03
7	0.786 ± 0.067	0.329 ± 0.258	0.252 ± 0.262	0.976 ± 0.069	0.786 ± 0.067
8	0.769 ± 0.059	0.276 ± 0.24	0.236 ± 0.273	0.96 ± 0.099	0.769 ± 0.059
9	0.765 ± 0.052	0.271 ± 0.225	0.234 ± 0.232	0.954 ± 0.074	0.765 ± 0.052
10	0.764 ± 0.049	0.255 ± 0.223	0.206 ± 0.193	0.963 ± 0.069	0.764 ± 0.049
11	0.762 ± 0.049	0.257 ± 0.212	0.216 ± 0.195	0.957 ± 0.081	0.762 ± 0.049
12	0.768 ± 0.078	0.282 ± 0.275	0.274 ± 0.249	0.945 ± 0.081	0.768 ± 0.078
13	0.785 ± 0.058	0.356 ± 0.244	0.336 ± 0.27	0.945 ± 0.083	0.785 ± 0.058
14	0.791 ± 0.063	0.368 ± 0.258	0.33 ± 0.26	0.956 ± 0.065	0.791 ± 0.063
15	0.792 ± 0.06	0.38 ± 0.237	0.34 ± 0.266	0.953 ± 0.07	0.792 ± 0.06
16	0.778 ± 0.066	0.313 ± 0.261	0.276 ± 0.253	0.958 ± 0.08	0.778 ± 0.066
17	0.785 ± 0.068	0.348 ± 0.268	0.328 ± 0.309	0.949 ± 0.079	0.785 ± 0.068
18	0.779 ± 0.047	0.31 ± 0.226	0.212 ± 0.174	0.981 ± 0.036	0.779 ± 0.047
19	0.792 ± 0.06	0.351 ± 0.256	0.254 ± 0.228	0.984 ± 0.036	0.792 ± 0.06
20	0.792 ± 0.054	0.36 ± 0.241	0.286 ± 0.23	0.972 ± 0.04	0.792 ± 0.054
21	0.792 ± 0.045	0.363 ± 0.214	0.252 ± 0.197	0.984 ± 0.039	0.792 ± 0.045
22	0.785 ± 0.05	0.336 ± 0.231	0.244 ± 0.209	0.978 ± 0.046	0.785 ± 0.05
23	0.794 ± 0.055	0.372 ± 0.237	0.296 ± 0.237	0.972 ± 0.045	0.794 ± 0.055
24	0.788 ± 0.048	0.364 ± 0.212	0.282 ± 0.214	0.969 ± 0.053	0.788 ± 0.048
25	**0.803 ± 0.071**	0.395 ± 0.27	0.33 ± 0.263	0.971 ± 0.049	**0.803 ± 0.071**
26	0.779 ± 0.054	0.331 ± 0.228	0.288 ± 0.229	0.955 ± 0.056	0.779 ± 0.054
27	0.781 ± 0.064	0.341 ± 0.247	0.298 ± 0.224	0.953 ± 0.068	0.781 ± 0.064
28	0.78 ± 0.064	0.347 ± 0.233	0.312 ± 0.214	0.947 ± 0.062	0.78 ± 0.064
29	0.783 ± 0.058	0.347 ± 0.234	0.308 ± 0.22	0.953 ± 0.054	0.783 ± 0.058
30	0.799 ± 0.06	**0.421 ± 0.221**	**0.368 ± 0.233**	0.954 ± 0.063	0.799 ± 0.06

TABLE C.4: The *UFS+RFE* feature performance measures for the **Full** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
2	0.831 ± 0.071	0.569 ± 0.184	0.664 ± 0.204	0.89 ± 0.079	0.831 ± 0.071
3	0.797 ± 0.079	0.483 ± 0.212	0.594 ± 0.213	0.87 ± 0.091	0.797 ± 0.079
4	0.826 ± 0.076	0.565 ± 0.196	0.688 ± 0.201	0.875 ± 0.078	0.826 ± 0.076
5	0.829 ± 0.079	0.58 ± 0.196	0.704 ± 0.171	0.874 ± 0.086	0.829 ± 0.079
6	0.837 ± 0.081	0.596 ± 0.212	0.708 ± 0.211	0.884 ± 0.085	0.837 ± 0.081
7	**0.866 ± 0.075**	**0.68 ± 0.179**	**0.796 ± 0.183**	0.891 ± 0.087	**0.866 ± 0.075**
8	0.839 ± 0.078	0.603 ± 0.197	0.728 ± 0.205	0.879 ± 0.08	0.839 ± 0.078
9	0.846 ± 0.073	0.623 ± 0.183	0.744 ± 0.194	0.882 ± 0.079	0.846 ± 0.073
10	0.849 ± 0.075	0.62 ± 0.202	0.72 ± 0.192	0.896 ± 0.075	0.849 ± 0.075
11	0.834 ± 0.068	0.592 ± 0.16	0.708 ± 0.156	0.879 ± 0.087	0.834 ± 0.068
12	0.853 ± 0.067	0.635 ± 0.17	0.734 ± 0.181	0.895 ± 0.077	0.853 ± 0.067
13	0.855 ± 0.075	0.653 ± 0.193	0.762 ± 0.222	0.889 ± 0.101	0.855 ± 0.075
14	0.856 ± 0.069	0.644 ± 0.182	0.74 ± 0.203	0.898 ± 0.074	0.856 ± 0.069
15	**0.866 ± 0.075**	0.673 ± 0.18	0.752 ± 0.19	0.907 ± 0.09	**0.866 ± 0.075**
16	**0.866 ± 0.07**	0.663 ± 0.169	0.712 ± 0.186	**0.921 ± 0.083**	**0.866 ± 0.07**
17	0.836 ± 0.073	0.582 ± 0.192	0.674 ± 0.201	0.894 ± 0.082	0.836 ± 0.073
18	0.845 ± 0.074	0.613 ± 0.191	0.714 ± 0.188	0.891 ± 0.08	0.845 ± 0.074
19	0.853 ± 0.067	0.633 ± 0.178	0.724 ± 0.202	0.899 ± 0.076	0.853 ± 0.067
20	0.861 ± 0.081	0.671 ± 0.183	0.78 ± 0.173	0.89 ± 0.098	0.861 ± 0.081
21	0.851 ± 0.072	0.635 ± 0.174	0.738 ± 0.183	0.891 ± 0.084	0.851 ± 0.072
22	0.84 ± 0.078	0.6 ± 0.207	0.706 ± 0.213	0.888 ± 0.079	0.84 ± 0.078
23	0.837 ± 0.083	0.609 ± 0.201	0.748 ± 0.197	0.869 ± 0.092	0.837 ± 0.083
24	0.828 ± 0.077	0.587 ± 0.182	0.73 ± 0.186	0.863 ± 0.091	0.828 ± 0.077
25	0.838 ± 0.075	0.604 ± 0.194	0.724 ± 0.215	0.879 ± 0.088	0.838 ± 0.075
26	0.837 ± 0.07	0.592 ± 0.184	0.706 ± 0.189	0.884 ± 0.073	0.837 ± 0.07
27	0.841 ± 0.068	0.62 ± 0.16	0.758 ± 0.175	0.871 ± 0.084	0.841 ± 0.068
28	0.833 ± 0.078	0.603 ± 0.183	0.746 ± 0.183	0.864 ± 0.094	0.833 ± 0.078
29	0.838 ± 0.08	0.61 ± 0.182	0.724 ± 0.178	0.879 ± 0.101	0.838 ± 0.08
30	0.842 ± 0.07	0.614 ± 0.173	0.718 ± 0.192	0.886 ± 0.088	0.842 ± 0.07

TABLE C.5: The *RF* features performance measures for the **Full** dataset.

C.2 Feature Sets Rankings

R	Univariate feature selection	Recursive feature elimination
1	SOFA_T2	APACHE II_T1
2	APACHE II_T1	LVOT average velocity time integral cm _T1
3	APACHE II_T2	HCO3 mmol L _T1
4	SOFA_T1	SOFA_T1
5	APACHE II_T3	Heart rate bpm _T3 1
6	Glasgow Coma Scale_T2	Weight kg
7	FiO2_T3	Na mmol L _T1
8	FiO2_T2	Leukocytes total 1000 mm 3 _T2
9	pH_T2	Mean arterial pressure mmHg _T3
10	Respiratory rate_T1	Sat O2 FiO2_T3
11	Sedation Scale SAS _T2	Diastolic Blood Pressure mmHg _T3
12	Tracheal Intubation_T2=Yes	Hematocrit_T2
13	Tracheal Intubation_T2=No	PaO2 mmHg _T2
14	SOFA_T3	E wave cm s _T1
15	Lactate levels mmol L _T2	Systolic blood pressure mmHg _T3
16	X Norepinephrine xb5g kg min _T1	Lympho abs count_T2
17	Sat O2 FiO2_T3	Tidal volume VT _T1
18	Sat O2 FiO2_T2	Heart rate bpm _T1 1
19	Number of affected organs	SvcO2_T1
20	Heart rate bpm _T3 1	APACHE II_T3
21	Tidal volume VT _T1	Chloride_T1
22	LA dilatation by eyeballing_T1=No	PT_T1
23	LA dilatation by eyeballing_T1=Yes	Glasgow Coma Scale_T2
24	Glasgow Coma Scale_T3	Systolic blood pressure mmHg _T2
25	Creatinine mg dL _T2	Platelet count_T1
26	Glasgow Coma Scale_T1	Base Excess mmol L _T1
27	Fluid Balance ml _T1	APACHE II_T2
28	Platelet count_T1	Value _T1
29	Creatinine mg dL _T3	PaCO2 mmHg _T1
30	K mmol L _T2	Glasgow Coma Scale_T1

TABLE C.6: The *UFS* and *RFE* feature rankings for the **Full** dataset.

```
PRECISION: APACHE II_T1, Diastolic Blood Pressure mmHg _T3, HCO3 mmol
   L_T1, Heart rate bpm _T3 1, Hematocrit _T2, LVOT average velocity
   time integral cm _T1, Leukocytes total 1000 mm 3 _T2, Mean arterial
    pressure mmHg _T3, Na mmol L _T1, SOFA_T1, Sat O2 FiO2_T3, Weight
   kg;
```

```
AUC: APACHE II_T1, APACHE II_T3, Chloride_T1, Diastolic Blood Pressure
    mmHg _T3, E wave cm s _T1, Glasgow Coma Scale_T2, HCO3 mmol L _T1,
    Heart rate bpm _T1 1, Heart rate bpm _T3 1, Hematocrit _T2, LVOT
    average velocity time integral cm _T1, Leukocytes total 1000 mm 3
    _T2, Lympho abs count_T2, Mean arterial pressure mmHg _T3, Na mmol
    L _T1, PT_T1, PaO2 mmHg _T2, SOFA_T1, Sat O2 FiO2_T3, SvcO2 _T1,
    Systolic blood pressure mmHg _T3, Tidal volume VT _T1, Weight kg;
```

List of features for the *RFECV* with the *precision* and the *AUC* metrics for the **Full** dataset.

R	UFS+FSE	Random Forest
1	LV Dilatation LVEDV _T1=Moderate	SOFA_T2
2	Aortic Valve Stenosis_T1=Moderate	APACHE II_T1
3	Is the patient under other drugs_-T1=Yes	APACHE II_T3
4	Is the patient under other drugs_-T1=No	Lactate levels mmol L _T2
5	Sedation Scale SAS _T2	Urine Output mL day _T1
6	SOFA_T1	Heart rate bpm _T3 1
7	Site of sampling=Urine	Respiratory rate_T1
8	SOFA_T3	APACHE II_T2
9	Glasgow Coma Scale_T3	Urine Output mL day _T2
10	Acute Kidney Injury_T2=AKIN III	X Norepinephrine_T1
11	APACHE II_T3	E wave_T1
12	Sedation Scale SAS _T1	FiO2_T3
13	HCO3 mmol L _T1	Sat O2 FiO2_T3
14	Respiratory rate_T1	SOFA_T1
15	Is the patient under other drugs_-T3=No	SOFA_T3
16	Is the patient under other drugs_-T3=Yes	pH_T2
17	Hematocrit_T3	Creatinine mg dL _T2
18	E A ratio_T1=1 39	Fluid Balance ml _T2
19	LA dilatation by eyeballing_T1=Yes	CRP_Value mg L _T3
20	Number of affected organs	K mmol L _T3
21	LA dilatation by eyeballing_T1=No	Tidal volume VT _T1
22	Base Excess mmol L _T1	Glasgow Coma Scale_T3
23	K mmol L _T2	Fluid Balance ml _T1
24	LVOT average diameter mm _T1	E e _T1
25	Heart rate bpm _T3 1	Glasgow Coma Scale_T2
26	K mmol L _T3	Pplat_T2
27	Respiratory rate rpm _T1	Temperature C _T3
28	PT_T2	PT_T2
29	Pplat_T1	Platelet count_T2
30	PEEP_T1	Heart rate bpm _T1

TABLE C.7: The *UFS+RFE* and the *RF* feature rankings for the **Full** dataset.

Appendix D

Chapter 5: T1+T2 dataset

D.1 Feature Selection Results

F	Accuracy	MCC	Sensitivity	Specificity	AUC
1	0.791 ± 0.066	0.416 ± 0.213	0.448 ± 0.21	**0.914 ± 0.069**	0.791 ± 0.066
2	0.836 ± 0.08	0.582 ± 0.212	0.666 ± 0.208	0.897 ± 0.083	0.836 ± 0.08
3	0.815 ± 0.067	0.536 ± 0.174	0.662 ± 0.176	0.869 ± 0.078	0.815 ± 0.067
4	0.832 ± 0.073	0.595 ± 0.171	0.734 ± 0.155	0.866 ± 0.088	0.832 ± 0.073
5	0.793 ± 0.078	0.511 ± 0.174	0.696 ± 0.166	0.827 ± 0.098	0.793 ± 0.078
6	0.801 ± 0.073	0.486 ± 0.206	0.606 ± 0.214	0.87 ± 0.08	0.801 ± 0.073
7	0.791 ± 0.081	0.476 ± 0.203	0.604 ± 0.185	0.858 ± 0.097	0.791 ± 0.081
8	0.825 ± 0.07	0.552 ± 0.192	0.63 ± 0.201	0.895 ± 0.08	0.825 ± 0.07
9	0.815 ± 0.076	0.536 ± 0.201	0.65 ± 0.237	0.874 ± 0.092	0.815 ± 0.076
10	0.797 ± 0.096	0.523 ± 0.205	0.68 ± 0.202	0.839 ± 0.119	0.797 ± 0.096
11	0.784 ± 0.11	0.534 ± 0.212	0.776 ± 0.17	0.787 ± 0.132	0.784 ± 0.11
12	0.803 ± 0.093	0.54 ± 0.201	0.728 ± 0.187	0.829 ± 0.108	0.803 ± 0.093
13	0.827 ± 0.084	0.593 ± 0.187	0.748 ± 0.176	0.856 ± 0.099	0.827 ± 0.084
14	0.839 ± 0.087	0.621 ± 0.191	0.756 ± 0.169	0.869 ± 0.105	0.839 ± 0.087
15	0.807 ± 0.094	0.557 ± 0.203	0.746 ± 0.181	0.829 ± 0.109	0.807 ± 0.094
16	0.815 ± 0.09	0.586 ± 0.188	**0.796 ± 0.157**	0.822 ± 0.101	0.815 ± 0.09
17	0.804 ± 0.096	0.551 ± 0.207	0.746 ± 0.165	0.824 ± 0.107	0.804 ± 0.096
18	0.837 ± 0.075	0.606 ± 0.175	0.742 ± 0.161	0.871 ± 0.084	0.837 ± 0.075
19	0.807 ± 0.086	0.549 ± 0.193	0.722 ± 0.19	0.838 ± 0.106	0.807 ± 0.086
20	0.817 ± 0.086	0.585 ± 0.186	0.778 ± 0.167	0.831 ± 0.099	0.817 ± 0.086
21	0.796 ± 0.076	0.523 ± 0.173	0.712 ± 0.18	0.826 ± 0.095	0.796 ± 0.076
22	0.815 ± 0.078	0.556 ± 0.194	0.718 ± 0.194	0.849 ± 0.087	0.815 ± 0.078
23	0.809 ± 0.077	0.545 ± 0.19	0.716 ± 0.198	0.842 ± 0.093	0.809 ± 0.077
24	0.812 ± 0.077	0.559 ± 0.186	0.738 ± 0.191	0.839 ± 0.091	0.812 ± 0.077
25	0.837 ± 0.08	0.615 ± 0.205	0.772 ± 0.215	0.861 ± 0.087	0.837 ± 0.08
26	0.849 ± 0.072	**0.641 ± 0.169**	0.764 ± 0.168	0.88 ± 0.09	0.849 ± 0.072
27	0.845 ± 0.076	0.609 ± 0.205	0.704 ± 0.207	0.895 ± 0.077	0.845 ± 0.076
28	**0.855 ± 0.06**	0.64 ± 0.155	0.736 ± 0.165	0.898 ± 0.07	**0.855 ± 0.06**
29	0.853 ± 0.069	0.64 ± 0.167	0.726 ± 0.176	0.899 ± 0.088	0.853 ± 0.069
30	0.839 ± 0.07	0.605 ± 0.181	0.716 ± 0.196	0.884 ± 0.084	0.839 ± 0.07

TABLE D.1: The *UFS* features performance measures for the **T1+T2** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
1	0.744 ± 0.052	0.186 ± 0.224	0.178 ± 0.141	**0.946 ± 0.057**	0.744 ± 0.052
2	0.772 ± 0.08	0.356 ± 0.245	0.392 ± 0.204	0.908 ± 0.083	0.772 ± 0.08
3	0.769 ± 0.078	0.398 ± 0.204	0.51 ± 0.201	0.862 ± 0.092	0.769 ± 0.078
4	0.681 ± 0.098	0.318 ± 0.19	0.65 ± 0.173	0.691 ± 0.125	0.681 ± 0.098
5	0.682 ± 0.096	0.311 ± 0.191	0.63 ± 0.209	0.701 ± 0.135	0.682 ± 0.096
6	0.694 ± 0.086	0.334 ± 0.179	0.646 ± 0.189	0.711 ± 0.117	0.694 ± 0.086
7	0.734 ± 0.081	0.379 ± 0.174	0.618 ± 0.17	0.775 ± 0.106	0.734 ± 0.081
8	0.725 ± 0.095	0.359 ± 0.221	0.604 ± 0.208	0.768 ± 0.118	0.725 ± 0.095
9	0.743 ± 0.084	0.383 ± 0.207	0.592 ± 0.213	0.797 ± 0.104	0.743 ± 0.084
10	0.727 ± 0.082	0.347 ± 0.218	0.578 ± 0.238	0.781 ± 0.103	0.727 ± 0.082
11	0.741 ± 0.097	0.392 ± 0.222	0.614 ± 0.214	0.786 ± 0.12	0.741 ± 0.097
12	0.732 ± 0.093	0.359 ± 0.225	0.58 ± 0.231	0.786 ± 0.114	0.732 ± 0.093
13	0.735 ± 0.085	0.384 ± 0.198	0.626 ± 0.201	0.774 ± 0.111	0.735 ± 0.085
14	0.726 ± 0.1	0.342 ± 0.229	0.562 ± 0.203	0.784 ± 0.121	0.726 ± 0.1
15	0.765 ± 0.076	0.403 ± 0.21	0.55 ± 0.239	0.841 ± 0.103	0.765 ± 0.076
16	0.769 ± 0.08	0.465 ± 0.179	0.678 ± 0.204	0.801 ± 0.111	0.769 ± 0.08
17	0.755 ± 0.077	0.385 ± 0.2	0.552 ± 0.21	0.828 ± 0.097	0.755 ± 0.077
18	0.764 ± 0.093	0.431 ± 0.223	0.626 ± 0.218	0.813 ± 0.103	0.764 ± 0.093
19	0.766 ± 0.072	0.436 ± 0.167	0.63 ± 0.18	0.814 ± 0.094	0.766 ± 0.072
20	0.782 ± 0.081	0.481 ± 0.197	0.674 ± 0.209	0.82 ± 0.098	0.782 ± 0.081
21	0.797 ± 0.08	0.521 ± 0.2	0.714 ± 0.204	0.826 ± 0.093	0.797 ± 0.08
22	0.785 ± 0.087	0.492 ± 0.214	0.686 ± 0.214	0.821 ± 0.105	0.785 ± 0.087
23	0.777 ± 0.081	0.479 ± 0.192	0.688 ± 0.205	0.809 ± 0.1	0.777 ± 0.081
24	0.791 ± 0.087	0.515 ± 0.203	0.72 ± 0.202	0.816 ± 0.102	0.791 ± 0.087
25	0.79 ± 0.089	0.487 ± 0.233	0.658 ± 0.23	0.837 ± 0.092	0.79 ± 0.089
26	0.813 ± 0.077	0.56 ± 0.196	0.74 ± 0.214	0.839 ± 0.09	0.813 ± 0.077
27	**0.819 ± 0.085**	**0.561 ± 0.212**	0.71 ± 0.205	0.858 ± 0.088	**0.819 ± 0.085**
28	0.813 ± 0.083	0.543 ± 0.214	0.69 ± 0.218	0.856 ± 0.093	0.813 ± 0.083
29	0.812 ± 0.077	0.551 ± 0.189	**0.714 ± 0.206**	0.847 ± 0.091	0.812 ± 0.077
30	0.806 ± 0.087	0.519 ± 0.233	0.664 ± 0.23	0.857 ± 0.089	0.806 ± 0.087

TABLE D.2: The *RFE* features performance measures for the **T1+T2** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
2	0.762 ± 0.078	0.313 ± 0.256	0.356 ± 0.205	**0.907 ± 0.076**	0.762 ± 0.078
3	0.757 ± 0.083	0.363 ± 0.214	0.482 ± 0.202	0.856 ± 0.097	0.757 ± 0.083
41	**0.812 ± 0.078**	**0.534 ± 0.201**	**0.662 ± 0.215**	0.866 ± 0.093	**0.812 ± 0.078**
42	0.808 ± 0.075	0.526 ± 0.181	0.658 ± 0.19	0.861 ± 0.09	0.808 ± 0.075
55	0.77 ± 0.086	0.444 ± 0.226	0.64 ± 0.237	0.816 ± 0.103	0.77 ± 0.086

TABLE D.3: The *RFECV* features performance measures for the **T1+T2** dataset. The first entry uses the *log loss*, followed by the *accuracy*, the *precision*, the *average precision* and the *f1-macro* metrics.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
1	0.72 ± 0.137	0.158 ± 0.193	0.16 ± 0.265	0.92 ± 0.271	0.72 ± 0.137
2	0.779 ± 0.024	0.313 ± 0.169	0.162 ± 0.092	**1.0 ± 0.0**	0.779 ± 0.024
3	0.766 ± 0.036	0.239 ± 0.208	0.154 ± 0.129	0.985 ± 0.029	0.766 ± 0.036
4	0.762 ± 0.034	0.215 ± 0.2	0.146 ± 0.129	0.981 ± 0.031	0.762 ± 0.034
5	0.783 ± 0.038	0.327 ± 0.196	0.226 ± 0.149	0.981 ± 0.031	0.783 ± 0.038
6	0.786 ± 0.037	0.349 ± 0.182	0.214 ± 0.13	0.99 ± 0.025	0.786 ± 0.037
7	0.784 ± 0.033	0.332 ± 0.183	0.208 ± 0.135	0.989 ± 0.026	0.784 ± 0.033
8	0.779 ± 0.039	0.307 ± 0.203	0.202 ± 0.143	0.985 ± 0.029	0.779 ± 0.039
9	0.784 ± 0.037	0.341 ± 0.187	0.22 ± 0.131	0.986 ± 0.029	0.784 ± 0.037
10	0.78 ± 0.041	0.317 ± 0.201	0.218 ± 0.144	0.981 ± 0.032	0.78 ± 0.041
11	0.776 ± 0.037	0.295 ± 0.196	0.192 ± 0.135	0.984 ± 0.03	0.776 ± 0.037
12	0.775 ± 0.04	0.286 ± 0.21	0.198 ± 0.154	0.981 ± 0.032	0.775 ± 0.04
13	0.774 ± 0.037	0.29 ± 0.193	0.188 ± 0.132	0.983 ± 0.031	0.774 ± 0.037
14	0.784 ± 0.035	0.334 ± 0.184	0.222 ± 0.149	0.984 ± 0.03	0.784 ± 0.035
15	0.773 ± 0.036	0.279 ± 0.197	0.2 ± 0.15	0.977 ± 0.033	0.773 ± 0.036
16	0.777 ± 0.038	0.296 ± 0.205	0.196 ± 0.147	0.984 ± 0.03	0.777 ± 0.038
17	0.775 ± 0.038	0.294 ± 0.199	0.192 ± 0.132	0.984 ± 0.03	0.775 ± 0.038
18	0.773 ± 0.037	0.301 ± 0.18	0.236 ± 0.151	0.964 ± 0.041	0.773 ± 0.037
19	0.776 ± 0.04	0.313 ± 0.188	0.238 ± 0.154	0.968 ± 0.043	0.776 ± 0.04
20	0.78 ± 0.036	0.331 ± 0.171	0.254 ± 0.147	0.968 ± 0.038	0.78 ± 0.036
21	0.769 ± 0.041	0.292 ± 0.184	0.232 ± 0.143	0.961 ± 0.046	0.769 ± 0.041
22	0.776 ± 0.035	0.319 ± 0.174	0.244 ± 0.149	0.966 ± 0.043	0.776 ± 0.035
23	0.769 ± 0.042	0.276 ± 0.202	0.21 ± 0.148	0.969 ± 0.044	0.769 ± 0.042
24	0.77 ± 0.041	0.288 ± 0.196	0.232 ± 0.152	0.962 ± 0.045	0.77 ± 0.041
25	0.776 ± 0.046	0.311 ± 0.209	0.232 ± 0.141	0.97 ± 0.042	0.776 ± 0.046
26	0.768 ± 0.044	0.265 ± 0.22	0.2 ± 0.162	0.971 ± 0.04	0.768 ± 0.044
27	0.781 ± 0.038	0.34 ± 0.182	0.26 ± 0.175	0.967 ± 0.046	0.781 ± 0.038
28	**0.791 ± 0.052**	**0.371 ± 0.216**	**0.296 ± 0.182**	0.967 ± 0.045	**0.791 ± 0.052**
29	0.784 ± 0.051	0.332 ± 0.222	0.266 ± 0.188	0.969 ± 0.041	0.784 ± 0.051
30	0.781 ± 0.049	0.332 ± 0.216	0.272 ± 0.178	0.963 ± 0.043	0.781 ± 0.049

TABLE D.4: The *UFS+RFE* features performance measures for the **T1+T2** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
1	0.758 ± 0.041	0.172 ± 0.225	0.166 ± 0.245	**0.969 ± 0.059**	0.758 ± 0.041
2	0.826 ± 0.068	0.515 ± 0.214	0.508 ± 0.214	0.939 ± 0.06	0.826 ± 0.068
3	0.806 ± 0.071	0.488 ± 0.197	0.532 ± 0.21	0.904 ± 0.086	0.806 ± 0.071
4	0.855 ± 0.081	0.652 ± 0.197	0.77 ± 0.188	0.885 ± 0.091	0.855 ± 0.081
5	**0.874 ± 0.073**	**0.693 ± 0.181**	**0.784 ± 0.187**	0.906 ± 0.079	**0.874 ± 0.073**
6	0.826 ± 0.081	0.596 ± 0.182	0.764 ± 0.177	0.848 ± 0.098	0.826 ± 0.081
7	0.84 ± 0.07	0.597 ± 0.182	0.688 ± 0.201	0.894 ± 0.079	0.84 ± 0.07
8	0.862 ± 0.059	0.651 ± 0.159	0.706 ± 0.191	0.917 ± 0.075	0.862 ± 0.059
9	0.834 ± 0.083	0.592 ± 0.205	0.694 ± 0.197	0.884 ± 0.096	0.834 ± 0.083
10	0.845 ± 0.075	0.615 ± 0.186	0.716 ± 0.188	0.891 ± 0.084	0.845 ± 0.075
11	0.852 ± 0.073	0.628 ± 0.185	0.7 ± 0.187	0.906 ± 0.082	0.852 ± 0.073
12	0.865 ± 0.061	0.664 ± 0.157	0.748 ± 0.173	0.906 ± 0.072	0.865 ± 0.061
13	0.841 ± 0.071	0.604 ± 0.197	0.726 ± 0.197	0.882 ± 0.074	0.841 ± 0.071
14	0.832 ± 0.074	0.6 ± 0.173	0.746 ± 0.181	0.862 ± 0.095	0.832 ± 0.074
15	0.828 ± 0.069	0.58 ± 0.168	0.706 ± 0.171	0.871 ± 0.086	0.828 ± 0.069
16	0.838 ± 0.068	0.607 ± 0.163	0.718 ± 0.19	0.881 ± 0.093	0.838 ± 0.068
17	0.828 ± 0.074	0.568 ± 0.182	0.666 ± 0.179	0.886 ± 0.086	0.828 ± 0.074
18	0.832 ± 0.071	0.592 ± 0.167	0.706 ± 0.191	0.876 ± 0.096	0.832 ± 0.071
19	0.828 ± 0.067	0.587 ± 0.162	0.732 ± 0.184	0.862 ± 0.085	0.828 ± 0.067
20	0.816 ± 0.074	0.551 ± 0.179	0.696 ± 0.175	0.859 ± 0.087	0.816 ± 0.074
21	0.813 ± 0.066	0.549 ± 0.162	0.704 ± 0.187	0.851 ± 0.09	0.813 ± 0.066
22	0.806 ± 0.076	0.538 ± 0.184	0.712 ± 0.184	0.839 ± 0.088	0.806 ± 0.076
23	0.825 ± 0.079	0.579 ± 0.182	0.72 ± 0.194	0.862 ± 0.098	0.825 ± 0.079
24	0.815 ± 0.08	0.568 ± 0.186	0.73 ± 0.186	0.846 ± 0.102	0.815 ± 0.08
25	0.817 ± 0.076	0.534 ± 0.197	0.62 ± 0.187	0.888 ± 0.092	0.817 ± 0.076
26	0.818 ± 0.073	0.543 ± 0.19	0.642 ± 0.199	0.881 ± 0.089	0.818 ± 0.073
27	0.809 ± 0.073	0.506 ± 0.2	0.594 ± 0.227	0.886 ± 0.091	0.809 ± 0.073
28	0.827 ± 0.068	0.558 ± 0.176	0.63 ± 0.203	0.897 ± 0.086	0.827 ± 0.068
29	0.81 ± 0.073	0.494 ± 0.211	0.556 ± 0.224	0.901 ± 0.081	0.81 ± 0.073
30	0.809 ± 0.066	0.511 ± 0.177	0.594 ± 0.216	0.886 ± 0.087	0.809 ± 0.066

TABLE D.5: The *RF* features performance measures for the **T1+T2** dataset.

D.2 Feature Sets Rankings

R	Precision	Area under the ROC curve
1	SOFA_T2	LVOT average diameter mm _T1
2	APACHE II_T1	SOFA_T1
3	APACHE II_T2	HCO3 mmol L _T1
4	SOFA_T1	Glasgow Coma Scale_T1
5	Glasgow Coma Scale_T2	PEEP_T2
6	FiO2_T2	PT_T1
7	pH_T2	PT_T2
8	Respiratory rate_T1	Respiratory rate rpm _T1
9	Sedation Scale SAS _T2	Na mmol L _T1
10	Tracheal Intubation_T2=Yes	Hematocrit _T1
11	Tracheal Intubation_T2=No	Value _T1
12	Lactate levels mmol L _T2	Heart rate bpm _T1 1
13	X Norepinephrine mg kg min _T1	Platelet count_T1
14	Sat O2 FiO2_T2	SvcO2 _T1
15	Number of affected organs	Tidal volume VT _T1
16	Tidal volume VT _T1	APACHE II_T1
17	LA dilatation by eyeballing_T1=No	aPTT_T2
18	LA dilatation by eyeballing_T1=Yes	SOFA_T2
19	Creatinine mg dL _T2	Hematocrit _T2
20	Glasgow Coma Scale_T1	APACHE II_T2
21	Fluid Balance ml _T1	Sat O2 FiO2_T2
22	Platelet count_T1	Respiratory rate_T2
23	K mmol L _T2	PaO2 FiO2_T2
24	Platelet count_T2	Base Excess mmol L _T1
25	LV Dilatation LVEDV _T1=Moderate	Platelet count_T2
26	E A ratio_T1=1 39	BMI
27	Aortic Valve Stenosis_T1=Moderate	Weight kg
28	Base Excess mmol L _T1	WBC abs count_T1
29	RBC count_T2	Base Excess mmol L _T2
30	Urine Output mL day _T1	PaO2 mmHg _T1

TABLE D.6: The *UFS* and the *RFE* feature rankings for the **T1+T2** dataset.

```
NEG LOG LOSS: LVOT average diameter mm _T1, SOFA_T1;

ACCURACY: HCO3 mmol L _T1, LVOT average diameter mm _T1, SOFA_T1;

PRECISION: APACHE II_T1, APACHE II_T2, BMI, Base Excess mmol L _T1,
    Base Excess mmol L _T2, CRP_Value mg L _T2, Diastolic Blood
    Pressure mmHg _T2, E wave cm s_T1, Glasgow Coma Scale _T1, Glasgow
    Coma Scale_T2, Glycemia mg dL _T1, HCO3 mmol L _T1, HCO3 mmol L _T2
    , Heart rate bpm _T1 1, Heart rate bpm _T2 1, Hematocrit_T1,
    Hematocrit _T2, LVOT average diameter mm _T1, Leukocytes total 1000
     mm 3 _T2, Na mmol L _T1, PEEP_T1, PEEP_T2, PT_T1, PT_T2, PaCO2
    mmHg _T1, PaO2 FiO2_T2, PaO2 mmHg _T1, Platelet count_T1, Platelet
    count_T2, Respiratory rate rpm _T1, Respiratory rate_T2, SOFA_T1,
    SOFA _T2, Sat O2 FiO2_T2, SvcO2 _T1, Tidal volume VT _T1, Value_T1,
     WBC abs count_T1, WBC abs count_T2, Weight kg , aPTT_T2;

AVERAGE PRECISION: APACHE II_T1, APACHE II_T2, BMI, Base Excess mmol L
    _T1, Base Excess mmol L _T2, CRP_Value mg L _T2, Diastolic Blood
    Pressure mmHg _T2, E wave cm s _T1, Glasgow Coma Scale_T1, Glasgow
    Coma Scale_T2, Glycemia mg dL _T1, HCO3 mmol L _T1, HCO3 mmol L _T2
    , Heart rate bpm _T1 1, Heart rate bpm _T2 1, Hematocrit _T1,
    Hematocrit _T2, LVOT average diameter mm _T1, Leukocytes total 1000
     mm 3 _T2, Na mmol L _T1, PEEP_T1, PEEP_T2, PT_T1, PT_T2, PaCO2
    mmHg _T1, PaO2 FiO2_T2, PaO2 mmHg _T1, Platelet count_T1, Platelet
    count_T2, Respiratory rate rpm _T1, Respiratory rate_T2, SOFA_T1,
    SOFA_T2, Sat O2 FiO2_T2, SvcO2 _T1, Systolic blood pressure mmHg
    _T2, Tidal volume VT _T1, Value _T1, WBC abs count_T1, WBC abs
    count_T2, Weight kg , aPTT_T2;

F1 MACRO: A wave cm s _T1, APACHE II_T1, APACHE II_T2, BMI, Base Excess
     mmol L_T1, Base Excess mmol L _T2, CRP_Value mg L _T1, CRP_Value
    mg L _T2, Diastolic Blood Pressure mmHg _T2, E wave cm s _T1,
    Glasgow Coma Scale_T1, Glasgow Coma Scale_T2, Glycemia mg dL _T1,
    HCO3 mmol L _T1, HCO3 mmol L _T2, Heart rate bpm _T1 1, Heart rate
    bpm _T2 1, Height cm , Hematocrit _T1, Hematocrit _T2, LA
    dilatation by eyeballing_T1=No, LA dilatation by eyeballing_T1=Yes,
     LVOT average diameter mm_T1, LVOT average velocity time integral
    cm _T1, Lactate levels mmol L _T1, Leukocytes total 1000 mm 3 _T1,
    Leukocytes total 1000 mm 3 _T2, Midazolam mg kg min _T1, Na mmol L
    _T1, Na mmol L _T2, PEEP_T1, PEEP_T2, PT_T1, PT_T2, PaCO2 mmHg_T1,
    PaO2 FiO2_T2, PaO2 mmHg_T1, PaO2 mmHg _T2, Platelet count_T1,
    Platelet count_T2, Pplat_T2, Respiratory rate rpm _T1, Respiratory
    rate_T2, SOFA_T1, SOFA_T2, Sat O2 FiO2_T2, Sat O2 _T2, SvcO2 _T1,
    Systolic blood pressure mmHg _T2, Tidal volume VT _T1, Value _T1,
    WBC abs count_T1, WBC abs count_T2, Weight kg, aPTT_T2;
```

List of features for the *RFECV* with the *precision* and the *average precision* metrics for the **T1+T2** dataset.

R	UFS+FSE	Random Forest
1	LV Dilatation LVEDV _T1=Moderate	Lactate levels mmol L _T2
2	E A ratio_T1=1 39	SOFA_T2
3	E A ratio_T1=0 69	Respiratory rate_T1
4	E A ratio_T1=0 6	APACHE II_T1
5	Aortic Valve Stenosis_T1=Moderate	E wave cm s _T1
6	Is the patient under other drugs_-T1=Yes	APACHE II_T2
7	Sedation Scale SAS _T2	X Norepinephrine mg kg min _T1
8	Site of sampling=CSF	Urine Output mL day _T2
9	Number of affected organs	Urine Output mL day _T1
10	Is the patient under other drugs_-T1=No	SOFA_T1
11	Site of sampling=Urine	pH_T2
12	Aortic Valve Regurgitation_-T2=Moderate	Tidal volume VT _T1
13	Aortic Valve Regurgitation_-T1=Moderate	Glasgow Coma Scale_T2
14	Is the patient under sedation drugs_-T2=No	Fluid Balance ml _T1
15	K mmol L _T2	Fluid Balance ml _T2
16	HCO3 mmol L _T1	Creatinine mg dL _T2
17	Segments affected by eyeballing_T2	Pplat_T2
18	E A ratio_T1=0 76	Lactate levels mmol L _T1
19	Glasgow Coma Scale_T1	Platelet count_T2
20	Temperature C _T1	E e _T1
21	Aortic Valve Regurgitation_-T1=Severe	Sat O2 FiO2_T2
22	Is the patient under sedation drugs_-T2=Yes	PT_T2
23	Acute Kidney Injury_T2=AKIN III	Respiratory rate_T2
24	SOFA_T1	Heart rate bpm _T1
25	Tricuspid Valve Regurgitation_-T2=Mild	Lympho abs count_T2
26	Base Excess mmol L _T1	PEEP_T2
27	Anuria and or RRT_T1=No	PaO2 mmHg _T2
28	Anuria and or RRT_T1=Yes	Pplat_T1
29	Respiratory rate_T1	LVOT average velocity time integral cm _T1
30	pH_T2	FiO2_T2

TABLE D.7: The *UFS+RFE* and the *RF* feature rankings for the **T1+T2** dataset.

Appendix E

Chapter 5: T1 dataset

E.1 Feature Selection Results

F	Accuracy	MCC	Sensitivity	Specificity	AUC
1	0.781 ± 0.065	0.392 ± 0.196	0.426 ± 0.183	0.907 ± 0.072	0.781 ± 0.065
2	0.774 ± 0.089	0.502 ± 0.185	**0.746 ± 0.176**	0.784 ± 0.115	0.774 ± 0.089
3	0.815 ± 0.07	0.566 ± 0.166	0.742 ± 0.18	0.841 ± 0.086	0.815 ± 0.07
4	0.794 ± 0.077	0.462 ± 0.213	0.56 ± 0.223	0.877 ± 0.088	0.794 ± 0.077
5	0.804 ± 0.081	0.514 ± 0.218	0.644 ± 0.233	0.861 ± 0.097	0.804 ± 0.081
6	0.793 ± 0.081	0.485 ± 0.22	0.636 ± 0.227	0.849 ± 0.094	0.793 ± 0.081
7	0.819 ± 0.072	0.545 ± 0.187	0.654 ± 0.192	0.879 ± 0.087	0.819 ± 0.072
8	0.807 ± 0.081	0.515 ± 0.195	0.618 ± 0.181	0.874 ± 0.1	0.807 ± 0.081
9	0.771 ± 0.087	0.46 ± 0.195	0.66 ± 0.184	0.81 ± 0.107	0.771 ± 0.087
10	0.779 ± 0.074	0.463 ± 0.172	0.626 ± 0.167	0.834 ± 0.095	0.779 ± 0.074
11	0.81 ± 0.077	0.54 ± 0.186	0.688 ± 0.18	0.854 ± 0.093	0.81 ± 0.077
12	0.823 ± 0.07	0.551 ± 0.178	0.638 ± 0.193	0.889 ± 0.091	0.823 ± 0.07
13	0.85 ± 0.056	0.595 ± 0.176	0.574 ± 0.201	0.949 ± 0.055	0.85 ± 0.056
14	0.833 ± 0.06	0.53 ± 0.202	0.468 ± 0.23	0.964 ± 0.052	0.833 ± 0.06
15	0.833 ± 0.06	0.526 ± 0.203	0.446 ± 0.196	**0.971 ± 0.046**	0.833 ± 0.06
16	0.847 ± 0.062	0.584 ± 0.193	0.538 ± 0.199	0.958 ± 0.061	0.847 ± 0.062
17	0.809 ± 0.063	0.46 ± 0.216	0.44 ± 0.214	0.941 ± 0.058	0.809 ± 0.063
18	0.805 ± 0.062	0.427 ± 0.234	0.386 ± 0.21	0.954 ± 0.047	0.805 ± 0.062
19	0.831 ± 0.063	0.524 ± 0.201	0.46 ± 0.201	0.963 ± 0.056	0.831 ± 0.063
20	0.828 ± 0.066	0.515 ± 0.22	0.478 ± 0.224	0.954 ± 0.054	0.828 ± 0.066
21	0.827 ± 0.063	0.511 ± 0.216	0.47 ± 0.225	0.954 ± 0.052	0.827 ± 0.063
22	0.829 ± 0.066	0.522 ± 0.218	0.494 ± 0.214	0.949 ± 0.055	0.829 ± 0.066
23	0.846 ± 0.066	0.575 ± 0.207	0.542 ± 0.216	0.955 ± 0.048	0.846 ± 0.066
24	0.852 ± 0.072	0.595 ± 0.225	0.598 ± 0.231	0.942 ± 0.06	0.852 ± 0.072
25	0.859 ± 0.071	0.627 ± 0.205	0.642 ± 0.216	0.937 ± 0.064	0.859 ± 0.071
26	0.858 ± 0.063	0.638 ± 0.17	0.706 ± 0.184	0.912 ± 0.068	0.858 ± 0.063
27	0.846 ± 0.071	0.612 ± 0.177	0.686 ± 0.196	0.904 ± 0.082	0.846 ± 0.071
28	0.86 ± 0.065	0.641 ± 0.184	0.718 ± 0.19	0.911 ± 0.066	0.86 ± 0.065
29	0.838 ± 0.061	0.594 ± 0.16	0.694 ± 0.193	0.889 ± 0.075	0.838 ± 0.061
30	**0.867 ± 0.063**	**0.668 ± 0.158**	0.736 ± 0.181	0.914 ± 0.074	**0.867 ± 0.063**

TABLE E.1: The *UFS* features performance measures for the **T1** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
1	0.609 ± 0.21	0.0 ± 0.0	**0.27 ± 0.444**	0.73 ± 0.444	0.609 ± 0.21
2	0.614 ± 0.208	0.0 ± 0.0	0.26 ± 0.439	0.74 ± 0.439	0.614 ± 0.208
3	0.737 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	**1.0 ± 0.0**	0.737 ± 0.0
4	0.732 ± 0.026	0.016 ± 0.086	0.028 ± 0.075	0.983 ± 0.048	0.732 ± 0.026
5	0.748 ± 0.036	0.105 ± 0.199	0.114 ± 0.214	0.974 ± 0.064	0.748 ± 0.036
6	0.741 ± 0.05	0.09 ± 0.19	0.108 ± 0.199	0.967 ± 0.079	0.741 ± 0.05
7	0.754 ± 0.057	0.145 ± 0.238	0.15 ± 0.234	0.97 ± 0.073	0.754 ± 0.057
8	0.727 ± 0.047	0.07 ± 0.151	0.116 ± 0.196	0.946 ± 0.103	0.727 ± 0.047
9	0.734 ± 0.066	0.1 ± 0.225	0.16 ± 0.27	0.939 ± 0.106	0.734 ± 0.066
10	0.735 ± 0.067	0.098 ± 0.224	0.15 ± 0.249	0.944 ± 0.106	0.735 ± 0.067
11	0.738 ± 0.053	0.085 ± 0.219	0.134 ± 0.259	0.954 ± 0.087	0.738 ± 0.053
12	0.738 ± 0.071	0.088 ± 0.237	0.124 ± 0.228	0.957 ± 0.094	0.738 ± 0.071
13	0.758 ± 0.059	0.179 ± 0.271	0.222 ± 0.306	0.95 ± 0.075	0.758 ± 0.059
14	0.74 ± 0.049	0.103 ± 0.224	0.154 ± 0.278	0.949 ± 0.091	0.74 ± 0.049
15	0.738 ± 0.079	0.121 ± 0.259	0.182 ± 0.297	0.937 ± 0.117	0.738 ± 0.079
16	0.746 ± 0.062	0.13 ± 0.253	0.178 ± 0.294	0.949 ± 0.093	0.746 ± 0.062
17	0.747 ± 0.048	0.134 ± 0.226	0.174 ± 0.266	0.951 ± 0.084	0.747 ± 0.048
18	0.735 ± 0.059	0.112 ± 0.225	0.166 ± 0.283	0.938 ± 0.112	0.735 ± 0.059
19	0.746 ± 0.06	0.128 ± 0.261	0.17 ± 0.283	0.952 ± 0.077	0.746 ± 0.06
20	**0.764 ± 0.067**	0.197 ± 0.296	0.222 ± 0.316	0.957 ± 0.069	**0.764 ± 0.067**
21	0.752 ± 0.078	0.158 ± 0.278	0.17 ± 0.275	0.96 ± 0.095	0.752 ± 0.078
22	0.747 ± 0.048	0.116 ± 0.221	0.126 ± 0.233	0.969 ± 0.071	0.747 ± 0.048
23	0.741 ± 0.075	0.142 ± 0.25	0.188 ± 0.274	0.938 ± 0.107	0.741 ± 0.075
24	0.748 ± 0.077	0.179 ± 0.273	0.232 ± 0.308	0.932 ± 0.108	0.748 ± 0.077
25	0.749 ± 0.074	0.157 ± 0.271	0.188 ± 0.291	0.95 ± 0.097	0.749 ± 0.074
26	0.751 ± 0.07	0.177 ± 0.272	0.214 ± 0.314	0.943 ± 0.104	0.751 ± 0.07
27	0.747 ± 0.055	0.159 ± 0.247	0.196 ± 0.281	0.944 ± 0.083	0.747 ± 0.055
28	0.738 ± 0.075	0.136 ± 0.247	0.19 ± 0.283	0.934 ± 0.119	0.738 ± 0.075
29	0.755 ± 0.058	**0.207 ± 0.253**	0.24 ± 0.294	0.939 ± 0.093	0.755 ± 0.058
30	0.738 ± 0.071	0.134 ± 0.251	0.17 ± 0.254	0.941 ± 0.103	0.738 ± 0.071

TABLE E.2: The *RFE* features performance measures for the **T1** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
1	0.599 ± 0.215	0.0 ± 0.0	0.29 ± 0.454	0.71 ± 0.454	0.599 ± 0.215
2	0.614 ± 0.208	0.0 ± 0.0	0.26 ± 0.439	0.74 ± 0.439	0.614 ± 0.208
8	0.739 ± 0.041	0.093 ± 0.176	0.12 ± 0.202	**0.961 ± 0.08**	0.739 ± 0.041
31	**0.75 ± 0.069**	**0.203 ± 0.252**	0.24 ± 0.284	0.932 ± 0.106	**0.75 ± 0.069**
33	0.744 ± 0.072	0.183 ± 0.276	0.222 ± 0.276	0.93 ± 0.094	0.744 ± 0.072
154	0.681 ± 0.083	0.169 ± 0.222	**0.374 ± 0.197**	0.791 ± 0.101	0.681 ± 0.083

TABLE E.3: The *RFECV* features performance measures for the **T1** dataset. The first entry uses the *recall* metric, followed by the *average precision*, the *weighted f1* and the *AUC*.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
1	0.741 ± 0.113	0.209 ± 0.197	0.156 ± 0.216	0.95 ± 0.218	0.741 ± 0.113
2	0.757 ± 0.026	0.154 ± 0.192	0.078 ± 0.098	**1.0 ± 0.0**	0.757 ± 0.026
3	0.749 ± 0.033	0.132 ± 0.197	0.102 ± 0.131	0.98 ± 0.032	0.749 ± 0.033
4	0.767 ± 0.037	0.241 ± 0.21	0.16 ± 0.136	0.984 ± 0.03	0.767 ± 0.037
5	0.769 ± 0.032	0.252 ± 0.198	0.162 ± 0.132	0.986 ± 0.029	0.769 ± 0.032
6	0.769 ± 0.031	0.27 ± 0.184	0.162 ± 0.105	0.986 ± 0.028	0.769 ± 0.031
7	0.775 ± 0.035	0.289 ± 0.196	0.178 ± 0.126	0.988 ± 0.027	0.775 ± 0.035
8	0.772 ± 0.033	0.264 ± 0.204	0.164 ± 0.137	0.989 ± 0.027	0.772 ± 0.033
9	0.767 ± 0.035	0.24 ± 0.207	0.158 ± 0.139	0.985 ± 0.029	0.767 ± 0.035
10	0.773 ± 0.036	0.261 ± 0.208	0.178 ± 0.152	0.985 ± 0.029	0.773 ± 0.036
11	0.763 ± 0.068	0.255 ± 0.228	0.194 ± 0.175	0.966 ± 0.106	0.763 ± 0.068
12	0.778 ± 0.044	0.318 ± 0.196	0.216 ± 0.159	0.979 ± 0.055	0.778 ± 0.044
13	0.771 ± 0.057	0.29 ± 0.221	0.204 ± 0.15	0.974 ± 0.067	0.771 ± 0.057
14	0.759 ± 0.059	0.269 ± 0.218	0.238 ± 0.164	0.946 ± 0.071	0.759 ± 0.059
15	0.772 ± 0.061	0.313 ± 0.215	0.254 ± 0.16	0.957 ± 0.074	0.772 ± 0.061
16	0.768 ± 0.071	0.299 ± 0.245	0.246 ± 0.192	0.955 ± 0.091	0.768 ± 0.071
17	0.769 ± 0.056	0.297 ± 0.212	0.216 ± 0.149	0.967 ± 0.074	0.769 ± 0.056
18	0.778 ± 0.057	0.346 ± 0.2	0.26 ± 0.169	0.964 ± 0.075	0.778 ± 0.057
19	0.765 ± 0.064	0.292 ± 0.234	0.242 ± 0.199	0.952 ± 0.096	0.765 ± 0.064
20	0.777 ± 0.057	0.333 ± 0.197	0.252 ± 0.164	0.964 ± 0.077	0.777 ± 0.057
21	**0.791 ± 0.043**	0.357 ± 0.206	0.236 ± 0.156	0.989 ± 0.028	**0.791 ± 0.043**
22	0.782 ± 0.049	0.321 ± 0.226	0.22 ± 0.159	0.983 ± 0.041	0.782 ± 0.049
23	0.782 ± 0.045	0.319 ± 0.216	0.22 ± 0.161	0.982 ± 0.034	0.782 ± 0.045
24	0.787 ± 0.042	0.345 ± 0.201	0.222 ± 0.152	0.989 ± 0.031	0.787 ± 0.042
25	0.777 ± 0.046	0.301 ± 0.21	0.208 ± 0.17	0.981 ± 0.057	0.777 ± 0.046
26	0.784 ± 0.043	0.345 ± 0.192	0.238 ± 0.152	0.979 ± 0.043	0.784 ± 0.043
27	0.783 ± 0.059	0.345 ± 0.225	0.274 ± 0.193	0.964 ± 0.076	0.783 ± 0.059
28	0.766 ± 0.084	0.374 ± 0.223	0.404 ± 0.274	0.896 ± 0.147	0.766 ± 0.084
29	0.776 ± 0.072	**0.393 ± 0.198**	0.396 ± 0.235	0.912 ± 0.123	0.776 ± 0.072
30	0.776 ± 0.078	0.39 ± 0.238	**0.42 ± 0.271**	0.903 ± 0.128	0.776 ± 0.078

TABLE E.4: The *UFS+RFE* features performance measures for the **T1** dataset.

F	Accuracy	MCC	Sensitivity	Specificity	AUC
1	0.788 ± 0.069	0.409 ± 0.212	0.436 ± 0.203	**0.914 ± 0.08**	0.788 ± 0.069
2	0.826 ± 0.081	0.562 ± 0.198	0.654 ± 0.165	0.888 ± 0.087	0.826 ± 0.081
3	0.822 ± 0.068	0.539 ± 0.185	0.614 ± 0.188	0.896 ± 0.078	0.822 ± 0.068
4	**0.848 ± 0.078**	**0.621 ± 0.206**	0.72 ± 0.198	0.894 ± 0.076	**0.848 ± 0.078**
5	0.821 ± 0.071	0.545 ± 0.188	0.638 ± 0.191	0.886 ± 0.083	0.821 ± 0.071
6	0.809 ± 0.078	0.516 ± 0.199	0.624 ± 0.197	0.875 ± 0.088	0.809 ± 0.078
7	0.829 ± 0.067	0.594 ± 0.167	**0.752 ± 0.188**	0.857 ± 0.083	0.829 ± 0.067
8	0.834 ± 0.082	0.603 ± 0.194	0.728 ± 0.213	0.872 ± 0.101	0.834 ± 0.082
9	0.813 ± 0.075	0.545 ± 0.193	0.688 ± 0.216	0.858 ± 0.094	0.813 ± 0.075
10	0.821 ± 0.061	0.532 ± 0.174	0.606 ± 0.189	0.897 ± 0.073	0.821 ± 0.061
11	0.807 ± 0.065	0.497 ± 0.184	0.592 ± 0.198	0.884 ± 0.074	0.807 ± 0.065
12	0.834 ± 0.077	0.577 ± 0.203	0.66 ± 0.201	0.896 ± 0.084	0.834 ± 0.077
13	0.834 ± 0.077	0.573 ± 0.205	0.652 ± 0.207	0.899 ± 0.078	0.834 ± 0.077
14	0.805 ± 0.076	0.51 ± 0.195	0.622 ± 0.208	0.871 ± 0.091	0.805 ± 0.076
15	0.821 ± 0.073	0.531 ± 0.223	0.62 ± 0.236	0.892 ± 0.071	0.821 ± 0.073
16	0.812 ± 0.069	0.534 ± 0.178	0.646 ± 0.202	0.871 ± 0.095	0.812 ± 0.069
17	0.811 ± 0.071	0.518 ± 0.204	0.63 ± 0.225	0.876 ± 0.084	0.811 ± 0.071
18	0.823 ± 0.075	0.542 ± 0.215	0.638 ± 0.226	0.889 ± 0.078	0.823 ± 0.075
19	0.806 ± 0.071	0.504 ± 0.198	0.616 ± 0.234	0.874 ± 0.09	0.806 ± 0.071
20	0.818 ± 0.08	0.532 ± 0.208	0.622 ± 0.19	0.889 ± 0.081	0.818 ± 0.08
21	0.813 ± 0.078	0.526 ± 0.215	0.63 ± 0.218	0.878 ± 0.09	0.813 ± 0.078
22	0.794 ± 0.07	0.471 ± 0.184	0.59 ± 0.199	0.866 ± 0.082	0.794 ± 0.07
23	0.805 ± 0.076	0.481 ± 0.227	0.564 ± 0.239	0.891 ± 0.083	0.805 ± 0.076
24	0.786 ± 0.075	0.469 ± 0.195	0.612 ± 0.226	0.849 ± 0.101	0.786 ± 0.075
25	0.773 ± 0.084	0.376 ± 0.255	0.474 ± 0.239	0.88 ± 0.084	0.773 ± 0.084
26	0.777 ± 0.071	0.407 ± 0.205	0.516 ± 0.212	0.87 ± 0.081	0.777 ± 0.071
27	0.791 ± 0.073	0.444 ± 0.209	0.526 ± 0.218	0.886 ± 0.083	0.791 ± 0.073
28	0.792 ± 0.071	0.453 ± 0.192	0.532 ± 0.194	0.884 ± 0.084	0.792 ± 0.071
29	0.78 ± 0.083	0.427 ± 0.225	0.548 ± 0.226	0.863 ± 0.093	0.78 ± 0.083
30	0.782 ± 0.069	0.407 ± 0.216	0.474 ± 0.229	0.892 ± 0.081	0.782 ± 0.069

TABLE E.5: The *RF* features performance measures for the **T1** dataset.

E.2 Feature Sets Rankings

R	Precision	Area under the ROC curve
1	APACHE II_T1	E A ratio_T1=0 6
2	SOFA_T1	Aortic Valve Regurgitation_-T1=Moderate
3	Respiratory rate_T1	Is the patient under other drugs_-T1=Yes
4	X Norepinephrine mg kg min _T1	Sedation Scale SAS _T1
5	Number of affected organs	LA dilatation by eyeballing_T1=Yes
6	Tidal volume VT _T1	Is the patient under other drugs_-T1=No
7	LA dilatation by eyeballing_T1=No	Aortic Valve Regurgitation_T1=No valv dysfunc
8	LA dilatation by eyeballing_T1=Yes	Glasgow Coma Scale_T1
9	Glasgow Coma Scale_T1	LV Hypertrophy LV mass _-T1=Moderate
10	Fluid Balance ml _T1	HCO3 mmol L _T1
11	Platelet count_T1	Respiratory rate_T1
12	LV Dilatation LVEDV _T1=Moderate	LVOT average velocity time integral cm _T1
13	E A ratio_T1=1 39	SOFA_T1
14	Aortic Valve Stenosis_T1=Moderate	LA dilatation by eyeballing_T1=No
15	Base Excess mmol L _T1	LV Hypertrophy LV mass _T1=No
16	Urine Output mL day _T1	Sat O2 _T1
17	Anuria and or RRT_T1=Yes	Fibrinogen_T1 g L
18	Anuria and or RRT_T1=No	Pplat_T1
19	LVOT average diameter mm _T1	Base Excess mmol L _T1
20	Mode_T1=Volume controlled	PT_T1
21	Lactate levels mmol L _T1	Lateral e cm s _T1
22	Is the patient under other drugs_-T1=Yes	Shock diagnosed within 48 from admission=No
23	Is the patient under other drugs_-T1=No	Neutro abs count_T1
24	Hypotension,MAP or SBP decrease _T1=Yes	Number of affected organs
25	Site of sampling=Urine	Respiratory rate rpm _T1
26	Hypotension,MAP or SBP decrease _T1=No	Hematocrit _T1
27	Mode_T1=Pressure controlled	Heart rate bpm _T1 1
28	Sedation Scale SAS _T1	Midazolam mg kg min _T1
29	Temperature C _T1	Weight kg
30	HCO3 mmol L _T1	aPTT_T1

TABLE E.6: The *UFS* and the *RFE* feature rankings for the **T1** dataset.

```
RECALL: E A ratio_T1=0 6;

AVERAGE PRECISION: Aortic Valve Regurgitation_T1=Moderate, E A ratio_T1
    =0 6;

PRECISION: Aortic Valve Regurgitation_T1=Moderate, Aortic Valve
    Regurgitation _T1=No valvular dysfunction, E A ratio_T1=0 6,
    Glasgow Coma Scale_T1, Is the patient under other drugs_T1=No, Is
    the patient under other drugs_T1=Yes, LA dilatation by
    eyeballing_T1=Yes, Sedation Scale SAS _T1;

F1 WEIGHTED: Aortic Valve Regurgitation_T1=Moderate, Aortic Valve
    Regurgitation _T1=No valvular dysfunction, Base Excess mmol L _T1,
    E A ratio_T1=0 6, Fibrinogen_T1 g L , Glasgow Coma Scale_T1, HCO3
    mmol L _T1, Heart rate bpm _T1 1, Hematocrit _T1, Is the patient
    under other drugs_T1=No, Is the patient under other drugs_T1=Yes,
    LA dilatation by eyeballing_T1=No, LA dilatation by eyeballing _T1=
    Yes, LV Hypertrophy LV mass _T1=Moderate, LV Hypertrophy LV mass
    _T1=No, LVOT average velocity time integral cm _T1, Lateral e cm s
    _T1, Midazolam mg kg min _T1, Neutro abs count_T1, Number of
    affected organs, PT_T1, Pplat_T1, Respiratory rate rpm _T1,
    Respiratory rate_T1, SOFA_T1, Sat O2 _T1, Sedation Scale SAS _T1,
    Value _T1, Was shock diagnosed within 48 from hospital admission=No
    , Weight kg, aPTT_T1;

AUC: Aortic Valve Regurgitation_T1=Moderate, Aortic Valve
    Regurgitation_T1=No valvular dysfunction, Base Excess mmol L _T1, E
     A ratio_T1=0 6, Fibrinogen_T1 g L, Glasgow Coma Scale_T1, HCO3
    mmol L _T1, Heart rate bpm _T1 1, Hematocrit _T1, Is the patient
    under other drugs_T1=No, Is the patient under other drugs_T1=Yes,
    LA dilatation by eyeballing_T1=No, LA dilatation by eyeballing_T1=
    Yes, LV Dilatation LVEDV _T1=No, LV Hypertrophy LV mass _T1=
    Moderate, LV Hypertrophy LV mass _T1=No, LVOT average velocity time
     integral cm _T1, Lateral e cm s _T1, Midazolam mg kg min _T1,
    Neutro abs count_T1, Number of affected organs, PT_T1, Pplat_T1,
    Respiratory rate rpm _T1, Respiratory rate_T1, SOFA_T1, Sat O2 _T1,
     Sedation Scale SAS _T1, SvcO2 _T1, Value _T1, Was shock diagnosed
    within 48 from hospital admission=No, Weight kg , aPTT_T1;
```

```
NEG LOG LOSS: A wave cm s _T1, AHF_T1=No, AHF_T1=no, AHF_T1=yes, APACHE
    II_T1, Acute Kidney Injury_T1=AKIN I, Acute Kidney Injury_T1=AKIN
   II, Acute Kidney Injury_T1=No, Anuria and or RRT_T1=No, Anuria and
   or RRT_T1=Yes, Aortic Valve Regurgitation_T1=Mild, Aortic Valve
   Regurgitation_T1=Moderate, Aortic Valve Regurgitation_T1=No
   valvular assessment available, Aortic Valve Regurgitation_T1=No
   valvular dysfunction, Aortic Valve Regurgitation_T1=Severe, Aortic
   Valve Stenosis_T1=Moderate, Aortic Valve Stenosis_T1=No valvular
   assessment available, BMI, Base Excess mmol L _T1, Bilirubin mg dL
   _T1, CRP_Value mg L _T1, Chloride_T1, Creatinine mg dL _T1,
   Diastolic Blood Pressure mmHg _T1, E A ratio_T1=0 6, E A ratio_T1=0
    62, E A ratio_T1=0 69, E A ratio_T1=0 72, E A ratio_T1=0 73, E A
   ratio_T1=0 78, E A ratio_T1=0 81, E A ratio_T1=0 84, E A ratio_T1=0
    85, E A ratio_T1=0 89, E A ratio_T1=0 97, E A ratio_T1=1, E A
   ratio_T1=1 01, E A ratio_T1=1 06, E A ratio_T1=1 18, E A ratio_T1=1
    3, E A ratio_T1=1 39, E A ratio_T1=1 68, E A ratio_T1=9, E e _T1,
   E wave cm s _T1, FiO2_T1, Fibrinogen_T1 g L , Gender=Female, Gender
   =Male, Glasgow Coma Scale_T1, Glycemia mg dL _T1, HCO3 mmol L _T1,
   Heart rate bpm _T1, Heart rate bpm _T1 1, Height cm , Hematocrit
   _T1, Hypotension SBP 90 mmHg or MAP 70 mmHg or SBP decrease 40 mmHg
    _T1=No, Hypotension SBP 90 mmHg or MAP 70 mmHg or SBP decrease 40
   mmHg _T1=Yes, Is the patient affected by Acute Myocardial
   Infarction AMI _T1=No, Is the patient affected by Acute Myocardial
   Infarction AMI _T1=Yes, Is the patient affected by Prolonged
   Arrhythmias PA _T1=No, Is the patient affected by Prolonged
   Arrhythmias PA _T1=Yes, Is the patient under inotropic drugs_T1=No,
    Is the patient under inotropic drugs_T1=Yes, Is the patient under
   other drugs_T1=No, Is the patient under other drugs_T1=Yes, Is the
   patient under sedation drugs_T1=No, Is the patient under sedation
   drugs_T1=Yes, LA dilatation by eyeballing_T1=No, LA dilatation by
   eyeballing_T1=Yes, LV Dilatation LVEDV _T1=Mild, LV Dilatation
   LVEDV _T1=Moderate, LV Dilatation LVEDV _T1=No, LV Dilatation LVEDV
    _T1=Severe, LV Hypertrophy LV mass _T1=Mild, LV Hypertrophy LV
   mass _T1=Moderate, LV Hypertrophy LV mass _T1=No, LVOT average
   diameter mm _T1, LVOT average velocity time integral cm _T1,
   Lactate levels mmol L _T1, Lateral e cm s _T1, Leukocytes total
   1000 mm 3 _T1, Lympho abs count_T1, Mean arterial pressure mmHg _T1
   , Midazolam mg kg min _T1, Mitral Valve Regurgitation_T1=Mild,
   Mitral Valve Regurgitation_T1=Moderate, Mitral Valve
   Regurgitation_T1=No valvular assessment available, Mitral Valve
   Regurgitation_T1=No valvular dysfunction, Mode_T1=Full pressure
   support, ...
```

```
... Mode_T1=Pressure controlled, Mode_T1=Volume controlled, Na mmol L
    _T1, Neutro abs count_T1, Number of affected organs, PEEP_T1, PT_T1
    , PaO2 FiO2_T1, PaO2 mmHg _T1, Paradoxical interventricular septum
    movements_T1=Moderate, Paradoxical interventricular septum
    movements_T1=No, Paradoxical interventricular septum movements_T1=
    Severe, Platelet count_T1, Platelets 10 3 mm 3 _T1, Pplat_T1, RBC
    count_T1, RV cross section D shape_T1=No, RV cross section D
    shape_T1=Yes, Renal Replacement Therapy_T1=No, Renal Replacement
    Therapy_T1=Yes, Respiratory rate rpm _T1, Respiratory rate_T1,
    Result=Negative, Result=Positive, Rhythm_T1=Atrial fibrillation,
    Rhythm_T1=Sinus, SOFA_T1, Sample 1=Negative, Sample 1=Positive,
    Sample 2=Negative, Sample 2=Positive, Sample 3=Negative, Sample 3=
    Positive, Sat O2 FiO2_T1, Sat O2 _T1, Sedation Scale SAS _T1,
    Segmental LV wall kinetics_T1=Normal, Segmental LV wall kinetics_T1
    =Reduced, Segments affected by eyeballing_T1, Site of sampling 1=
    Abdominal drain, Site of sampling 1=Other, Site of sampling 1=
    Respiratory tract, Site of sampling 1=Wound cultures, Site of
    sampling 2=Respiratory tract, Site of sampling 2=Urine, Site of
    sampling=Abdominal drain, Site of sampling=CSF, Site of sampling=
    Other, Site of sampling=Respiratory tract, Site of sampling=Urine,
    Site of sampling=Wound cultures, SvcO2 _T1, Systolic blood pressure
     mmHg _T1, Temperature C _T1, The patient was already receiving
    antibiotic treatment and this coverage has been maintained until
    the onset of sepsis=No, The patient was already receiving
    antibiotic treatment and this coverage has been maintained until
    the onset of sepsis=Yes, Tidal volume VT _T1, Tracheal
    Intubation_T1=No, Tracheal Intubation_T1=Yes, Tricuspid Valve
    Regurgitation_T1=Mild, Tricuspid Valve Regurgitation_T1=Moderate,
    Tricuspid Valve Regurgitation_T1=No valvular assessment available,
    Tricuspid Valve Regurgitation_T1=No valvular dysfunction, Tricuspid
     annular plane systolic excursion TAPSE mm _T1, Urine Output mL day
     _T1, Value _T1, WBC abs count_T1, Was shock diagnosed within 48
    from hospital admission=No, Was shock diagnosed within 48 from
    hospital admission=Yes, Was the patient transfused_T1=No, Was the
    patient transfused_T1=Yes, Weight kg , Which type of shock=
    Cardiogenic, Which type of shock=Septic, X Norepinephrine mg kg min
     _T1, aPTT_T1, pH_T1;
```

List of features for the *RFECV* with the *average precision* and the *accuracy* metrics for the **T1** dataset.

R	UFS+FSE	Random Forest
1	E A ratio_T1=1 39	APACHE II_T1
2	Site of sampling=CSF	Urine Output mL day _T1
3	E A ratio_T1=0 69	Respiratory rate_T1
4	Aortic Valve Stenosis_T1=Moderate	E wave cm s _T1
5	Mode_T1=Volume controlled	Lactate levels mmol L _T1
6	LA dilatation by eyeballing_T1=Yes	Tidal volume VT _T1
7	RBC count_T1	SOFA_T1
8	Is the patient under other drugs_-T1=Yes	Fluid Balance ml _T1
9	Is the patient under sedation drugs_-T1=No	E e _T1
10	Is the patient under other drugs_-T1=No	X Norepinephrine mg kg min _T1
11	E A ratio_T1=0 81	Heart rate bpm _T1
12	Is the patient under sedation drugs_-T1=Yes	Pplat_T1
13	Glasgow Coma Scale_T1	Temperature C _T1
14	LV Hypertrophy LV mass _-T1=Moderate	pH_T1
15	SOFA_T1	Fibrinogen_T1 g L
16	Sedation Scale SAS _T1	Respiratory rate rpm _T1
17	Aortic Valve Regurgitation_T1=No val dysfunc	Platelet count_T1
18	HCO3 mmol L _T1	PEEP_T1
19	pH_T1	Leukocytes total 1000 mm 3 _T1
20	FiO2_T1	LVOT av vel time integral cm _T1
21	LV Dilatation LVEDV _T1=Moderate	PaO2 FiO2_T1
22	Aortic Valve Regurgitation_-T1=Severe	Midazolam mg kg min _T1
23	Respiratory rate_T1	Base Excess mmol L _T1
24	Is the patient under inotropic drugs_T1=No	Neutro abs count_T1
25	LA dilatation by eyeballing_T1=No	A wave cm s _T1
26	LV Dilatation LVEDV _T1=No	Glycemia mg dL _T1
27	Hypotension, MAP or SBP decrease _T1=Yes	Platelets 10 3 mm 3 _T1
28	Hypotension, MAP or SBP decrease _T1=No	RBC count_T1
29	Mean arterial pressure mmHg _T1	LVOT average diameter mm _T1
30	Is the patient under inotropic drugs_T1=Yes	Lympho abs count_T1

TABLE E.7: The *UFS+RFE* and the *RF* feature rankings for the **T1** dataset.

Appendix F

Chapter 5: Stability Scores

Full+		Random Forest	
Tricuspid.regurgitation.max.velT1	1	Inferior.vena.cava.dist.indx_T3	0.688
Platelet.count_T1	0.8	Lactate.levels..mmol.L._T2	0.584
K.Ur_T1	0.8	SOFA_T2	0.568
Tidal.volume.VT_T1	0.8	SOFA_T3	0.548
Diastolic.Blood.Pressure.T3	0.6	Na.Ur_T3	0.512
Neutro.abs.count_T2	0.6	APACHE.II_T2	0.508
APACHE.II_T1	0.6	FiO2_T3	0.492
APACHE.II_T2	0.6	APACHE.II_T1	0.456
Respiratory.rate.rpm_T1	0.6	PCT_Value..mg.mL._T1	0.454
E.wave.cms_T1	0.6	X.Norepinephrine_T2	0.436
Value_T1	0.6	K.Ur_T2	0.418
Hematocrit_T1	0.6	APACHE.II_T3	0.414
Tricuspid.regurgitation.max.velT2	0.6	Urine.Output..mL.day._T1	207
Creat.Ur_T1	0.6	Respiratory.rate_T1	0.404
PT_T1	0.6	Sat.O2.FiO2_T3	0.4
Creat.Ur_T3	0.6	Pplat_T3	0.398
aPTT_T1	0.6	X..Dobutamine_T3	0.396
HCO3.mmolL_T1	0.6	Lateral.e._T3	0.394
Tidal.volume.VT_T3	0.6	Tricuspid.regurgitation.max.velT3	0.394
Inferior.vena.cava.dist.indx_T3	0.4	Tidal.volume_T2	0.39
Number.of.affected.organs	0.4	Heart.rate..bpm._T3.1	0.386
APACHE.II_T3	0.4	E.e._T3	0.382
Respiratory.rate_T1	0.4	E.wave_T1	0.38
K.Ur_T3	0.4	X..Norepinephrine_T1	0.378
LAdilatation.by.eyeballing_-T1=Yes	0.4	Tricuspid.annular.plane.syst.exc._-T3	0.378
Tracheal.Intubation_T2=Yes	0.4	Na.Ur_T2	0.372
PCT_Value.mgmL_T2	0.4	CRP_Value_T3	0.372
Tracheal.Intubation_T2=No	0.4	Urine.Output._T2	0.366
CRP_Value.mgL_T3	0.4	Glasgow.Coma.Scale_T3	0.364
LAdilatation.by.eyeballing_-T1=No	0.4	SOFA_T1	0.356

TABLE F.1: The first 30 features stability scores using the **Full+** dataset: for the *RF* (Random Forest) feature selection and for the rest of the feature selection techniques (*Full+*).

Full		**Random Forest**	
Heart.rate.bpm_T3.1	1	SOFA_T2	0.686
SOFA_T1	1	APACHE.II_T1	0.678
APACHE.II_T1	0.8	APACHE.II_T3	0.624
APACHE.II_T3	0.8	Lactate.levels..mmol.L._T2	0.614
HCO3.mmol.L_T1	0.8	Urine.Output..mL.day._T1	0.614
SatO2.FiO2_T3	0.8	Heart.rate.bpm._T3.1	0.606
Diastolic.Blood.Pressure.mmHg_-T3	0.6	Respiratory.rate_T1	0.598
Mean.arterial.pressure.mmHg_-T3	0.6	APACHE.II_T2	0.596
Na.mmolL_T1	0.6	Urine.Output..mL.day._T2	0.554
Leukocytes.total.1000mm3_T2	0.6	X.Norepinephrine_T1	0.552
Hematocrit_T2	0.6	E.wave..cm.s._T1	0.542
Weight.kg	0.6	FiO2_T3	0.524
Glasgow.Coma.Scale_T2	0.6	Sat.O2.FiO2_T3	0.52
Tidal.volume.VT_T1	0.6	SOFA_T1	0.516
Platelet.count_T1	0.2	SOFA_T3	0.51
APACHE.II_T2	0.2	pH_T2	0.508
K.mmolL_T2	0.2	Creatinine..mg.dL._T2	0.5
Sv.cO2_T1	0.2	Fluid.Balance..ml._T2	0.496
E.wave.cms_T1	0.2	CRP_Value..mg.L._T3	0.49
Heart.rate.bpm_T11	0.2	K..mmol.L._T3	0.486
Respiratory.rate_T1	0.2	Tidal.volume..VT._T1	0.486
Number.of.affected.organs	0.2	Glasgow.Coma.Scale_T3	0.476
LAdilatation.by.eyeballing_-T1=No	0.2	Fluid.Balance..ml._T1	0.474
PaO2mmHg_T2	0.2	E.e._T1	0.46
Chloride_T1	0.2	Glasgow.Coma.Scale_T2	0.452
Systolic.blood.pressure.mmHg_-T3	0.2	Pplat_T2	0.45
PT_T1	0.2	Temperature...C._T3	0.444
Glasgow.Coma.Scale_T1	0.2	PT_T2	0.434
Glasgow.Coma.Scale_T3	0.2	Platelet.count_T2	0.43
Lympho.abs.count_T2	0.2	Heart.rate..bpm._T1	0.426

TABLE F.2: The first 30 features stability scores using the **Full** dataset: for the *RF* (Random Forest) feature selection and for the rest of the feature selection techniques (*Full*).

T1+T2		**Random Forest**	
SOFA_T1	1.0	Lactate levels mmol L _T2	0.756
Base Excess mmol L _T1	0.75	SOFA_T2	0.748
HCO3 mmol L _T1	0.75	Respiratory rate_T1	0.734
LVOT average diameter mm _-T1	0.75	APACHE II_T1	0.72
Glasgow Coma Scale_T1	0.75	E wave cm s _T1	0.706
APACHE II_T2	0.625	APACHE II_T2	0.68
APACHE II_T1	0.625	X Norepinephrine mg kg min _T1	0.68
Platelet count_T1	0.625	Urine Output mL day _T2	0.668
Platelet count_T2	0.625	Urine Output mL day _T1	0.66
Sat O2 FiO2_T2	0.625	SOFA_T1	0.658
SOFA_T2	0.625	pH_T2	0.616
Tidal volume VT _T1	0.625	Tidal volume VT _T1	0.614
Weight kg	0.5	Glasgow Coma Scale_T2	0.612
PEEP_T2	0.5	Fluid Balance ml _T1	0.606
Na mmol L _T1	0.5	Fluid Balance ml _T2	0.604
SvcO2 _T1	0.5	Creatinine mg dL _T2	0.6
PaO2 mmHg _T1	0.5	Pplat_T2	0.58
aPTT_T2	0.5	Lactate levels mmol L _T1	0.564
Respiratory rate rpm _T1	0.5	Platelet count_T2	0.552
WBC abs count_T1	0.5	E e _T1	0.552
BMI	0.5	Sat O2 FiO2_T2	0.544
PT_T2	0.5	PT_T2	0.538
PT_T1	0.5	Respiratory rate_T2	0.538
Respiratory rate_T2	0.5	Heart rate bpm _T1	0.534
PaO2 FiO2_T2	0.5	Lympho abs count_T2	0.53
Heart rate bpm _T1 1	0.5	PEEP_T2	0.512
Base Excess mmol L _T2	0.5	PaO2 mmHg _T2	0.512
Hematocrit _T1	0.5	Pplat_T1	0.488
Hematocrit _T2	0.5	LVOT average velocity time integral cm _T1	0.484
Value _T1	0.5	FiO2_T2	0.476

TABLE F.3: The first 30 features stability scores using the **T1+T2** dataset: for the *RF* (Random Forest) feature selection and for the rest of the feature selection techniques (*T1+T2*).

T1		Random Forest	
LA dilatation by eyeballing_-T1=Yes	0.778	APACHE II_T1	0.898
E A ratio_T1=0.6	0.778	Urine Output mL day _T1	0.882
Sedation Scale SAS _T1	0.778	Respiratory rate_T1	0.878
Is the patient under other drugs_T1=No	0.778	E wave cm s _T1	0.844
Glasgow Coma Scale_T1	0.778	Lactate levels mmol L _T1	0.822
Is the patient under other drugs_T1=Yes	0.778	Tidal volume VT _T1	0.808
SOFA_T1	0.667	SOFA_T1	0.804
Respiratory rate_T1	0.667	Fluid Balance ml _T1	0.794
Aortic Valve Regurgitation_-T1=Moderate	0.667	E e _T1	0.786
Aortic Valve Regurg._T1=No val.dysfunc.	0.667	X Norepinephrine mg kg min _T1	0.782
HCO3 mmol L _T1	0.667	Heart rate bpm _T1	0.776
LA dilatation by eyeballing_-T1=No	0.667	Pplat_T1	0.764
Number of affected organs	0.556	Temperature C _T1	0.72
Base Excess mmol L _T1	0.556	pH_T1	0.706
LV Hypertrophy LV mass _-T1=Moderate	0.556	Fibrinogen_T1 g L	0.706
Weight kg	0.444	Respiratory rate rpm _T1	0.7
LV Hypertrophy LV mass _-T1=No	0.444	Platelet count_T1	0.68
Sat O2 _T1	0.444	PEEP_T1	0.676
Lateral e cm s _T1	0.444	Leukocytes total 1000 mm 3 _-T1	0.656
aPTT_T1	0.444	LVOT avg.vel.time integral_T1	0.656
Midazolam mg kg min _T1	0.444	PaO2 FiO2_T1	0.654
Fibrinogen_T1 g L	0.444	Midazolam mg kg min _T1	0.65
Respiratory rate rpm _T1	0.444	Base Excess mmol L _T1	0.64
LVOT avg. vel. time integral cm _T1	0.444	Neutro abs count_T1	0.64
PT_T1	0.444	A wave cm s _T1	0.632
Heart rate bpm _T1 1	0.444	Glycemia mg dL _T1	0.628
Neutro abs count_T1	0.444	Platelets 10 3 mm 3 _T1	0.628
Diagnosed within 48 from admission=No	0.444	RBC count_T1	0.622
Hematocrit _T1	0.444	LVOT average diameter mm _-T1	0.62
Pplat_T1	0.444	Lympho abs count_T1	0.616

TABLE F.4: The first 30 features stability scores using the **T1** dataset: for the *RF* (Random Forest) feature selection and for the rest of the feature selection techniques (*T1*).